ARMSTRONG LABORATORY

AL/CF-TR-1996-0093

ASSESSING THE PERFORMANCE IMPACT OF G-FORCES: DESIGN OF THE ACCELERATION-PERFORMANCE ASSESSMENT SIMULATION SYSTEM (A-PASS)

Robert D. O'Donnell
Rebecca Cardenas
Douglas Eddy

NTI, INCORPORATED
4130 LINDEN AVE.
SUITE 235
DAYTON, OHIO 45432

Robert Shaw

FCI, ASSOCIATES
1831 STONEWOOD DR.
DAYTON OH 45432

December 1995

FINAL REPORT FOR THE PERIOD MAY 1995 TO NOVEMBER 1995

Approved for public release; distribution is unlimited

AIR FORCE MATERIEL COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6573

NOTICES

DISCLAIMER

TECHNICAL REVIEW AND APPROVAL

AL/CF-TR-1996-0093

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE DIRECTOR

Albert S. Torigian

ALBERT S. TORIGIAN, Lt Colonel, USAF
Deputy
Biodynamics and Biocommunications Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1995	Final Report (25 May 95 - 15 Nov 95)

4. TITLE AND SUBTITLE
Assessing the Performance Impact of G-Forces: Design of the Acceleration Performance Assessment Simulation System

5. FUNDING NUMBERS
C: F41624-95-C-6005
PE: 65502F
PR: 3005
TA: CB
WU: 5C

6. AUTHOR(S)

Robert D. O'Donnell, Rebecca Cardenas, Douglas Eddy, Robert Shaw

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NTI, Inc.
4130 Linden Ave., Suite 235
Dayton OH 45432

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Armstrong Laboratory, Crew Systems Directorate
Biodynamics and Biocommunications Division
Human Systems Center
Air Force Materiel Command
Wright-Patterson AFB OH 45433-7008

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AL/CF-TR-1996-0093

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

A performance assessment system for use on a man-rated centrifuge is discussed. The problem of measuring human performance during high G flight simulation is addressed. This research addressed the fundamental problem of using laboratory data to assess the operational military impact (OMI) of physiological stresses. First, a battery of flight task simulations was conceptualized, based on current performance assessment theory. Secondy, a procedure was demonstrated for converting laboratory measures from this battery into measures of OMI. Specifically, performance data were entered into high-fidelity computer models of aircraft missions, yielding estimates of the military impact of stressors on human performance (e.g., kill probability, circular error - CEP, survivability, etc.). In the pilot demonstration, the effect of a hypothetical stressor on a pop-up air-to-ground maneuver in the F4E aircraft was evaluated. In the undegraded state, the CEP was 25.0 feet (S.D. 10.8 feet). In the degraded state, the CEP was 50.1 feet (S.D. 21.9 feet). This demonstration proved the feasibility of providing operationally meaningful metrics based on laboratory performance data.

14. SUBJECT TERMS
Flight Simulation, Performance Assessment, High G, Air Combat, Performance Model, Workload, Performance Metrics

15. NUMBER OF PAGES

70

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

NSN 7540-01-280-5500
Computer Generated 1996

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

THIS PAGE LEFT BLANK INTENTIONALLY.

PREFACE

The work supporting this report was carried out under a Phase I Small Business Innovation Research (SBIR) contract from the Armstrong Laboratory (Human Resources Division) to the performing organization. The contract technical monitor for this effort was Dr. Tamara Chelette (AL/CFBS).

The authors would like to thank the two pilot consultants who contributed to this effort: Mr. William Ercoline and Mr. Robert Shaw, of FCI Associates, Incorporated. The professional input of these experienced pilots was critical in all phases of the effort.

TABLE OF CONTENTS

Page

EXECUTIVE SUMMARY vi

1. INTRODUCTION. 1

1.1 BACKGROUND........ 1

2. OVERVIEW OF THE APPROACH TAKEN........ 5

3. OVERALL CONCEPT OF THE A-PASS PACKAGE........ 7

3.1 THE ISOLATION OF SKILLS IN A COMPLEX PERFORMANCE. 7
3.2 ACCOUNTING FOR INTERDEPENDENCIES AND VARIABILITY IN PERFORMANCE. 9

4. RESULTS OF PHASE I EFFORTS........ 12

4.1 OVERVIEW. 12
4.2 DETERMINATION OF OPERATIONAL "BOUNDARY" MISSIONS. 13
4.3 DEVELOPMENT OF THE A-PASS TEST BATTERY. 16
4.4 INTEGRATION OF COMPUTER MODELING TECHNIQUES. 49

5. DISCUSSION OF POTENTIAL PROBLEMS........ 57

6. REFERENCES........ 59

LIST OF ILLUSTRATIONS

Figure
Page

1 Schematic Representation of Skills Involved in the "Missile Defense" Maneuver 8

2 Perception of Relative Motion Test 18

3 The precision timing task. 23

4 The motion inference test. 27

5 The ILS approach task. 31

6 The pitch/roll capture test. 35

7 The gunsight tracking task. 37

8 The peripheral monitoring test. 40

9 The unusual attitudes test. 44

10 Micro Saint model of critical tasks in a pop-up air-to-ground attack. 54

11 Integrated Micro Saint/Systems modeling of the pop-up air-to-ground maneuver 56

EXECUTIVE SUMMARY

This effort arose out of an apparent dissatisfaction in the operational fighter pilot community with the types of data that were being generated in acceleration research. This dissatisfaction was symbolized by the comment of a fighter pilot after hearing the presentation of an excellent experimental study assessing performance in the acceleration environment. Even though appropriate and sensitive experimental measures were used, in a scientifically rigorous experimental design, this pilot essentially asked "what does this mean to me?" The implication was, of course, that no matter how well controlled and precise the study was, there was no way to interpret the operational impact of any results because the performance metric was phrased in laboratory terms.

The present effort therefore attempted to devise a defensible technique that would result in estimates of the actual Operational Military Impact (OMI) of an acceleration stress on the individual. It was recognized that a number of factors made this a formidable task. First, OMI is a general concept that relates to a vast number of possible military missions. Clearly, it would be impossible to study every mission. Secondly, there can be wide variation in human performance after any given stressor. Therefore, it would be impossible to develop a single number for OMI. Rather, a probability statement, with a range of variation, would be necessary. Third, if genuine precision in the OMI estimate was desired, it would be necessary to retain the strict scientific principals of experimentation. This requirement dictated that most data be generated in the laboratory centrifuge environment. In that environment, stimulus conditions and testing procedures can be rigorously controlled in ways that are simply not possible in actual flight, or even in many simulator situations. Finally, it became clear that performance metrics that had become routine, and which might be appropriate for laboratory studies in other environments, are not serving as adequate data sources for any measure of OMI. For instance, even the most sophisticated tracking measures were suspect in the operational community because of the lack of face validity. More critically, the range of measures available to the acceleration researcher (due to the constraints of the centrifuge environment) simply was not large enough to yield measures of all of the human resources that might be degraded under acceleration stress.

As a result of the above factors, it was recognized that there were three major developments which were required in the present program:

a. A method had to be developed to "bound" the number of operational missions or tasks that had to be studied in order to provide the pilot community with a full set of measures.

b. A suite of performance measures had to be developed which surveyed all of the important performance resources that might be degraded by acceleration stress.

c. A technique for translating performance measures into meaningful measures of OMI had to be developed.

The present Phase I SBIR effort addressed all three of these issues, providing preliminary solutions and demonstration of each. These efforts are described separately below.

THE USE OF "CRITICAL TASKS" TO BOUND THE PROBLEM.

With respect to bounding the number of missions that had to be tested, the concept of "critical tasks" was employed. Critical tasks refer to those piloting missions, mission segments, or other behaviors that are crucial to survival and mission accomplishment. It was hypothesized that experienced pilots could define a finite set of such critical tasks with which 1) were absolutely necessary for survival and/or mission accomplishment and 2) required all of the skills necessary for less critical tasks. In other words, the critical tasks represented a sub-set of all piloting behaviors. The tasks comprising this subset were demanding enough that, if the pilot could perform these tasks, it could safely be assumed that he or she could perform all other tasks in the cockpit.

This concept of "critical tasks" meant that one could now study only those tasks, with some assurance that performance measures on them would generalize to all cockpit tasks. To the degree that a pilot was decremented on any or all of these tasks, it could also be assumed that the pilot would be similarly decremented on less critical tasks, since all tasks would have required the same types of skills.

In the present effort, pilot consultants, in meetings with Government personnel, developed a list of nine such crucial tasks. These included: monitoring tasks under high workload, collision avoidance, missile avoidance, general timing ability, ILS landing, unusual attitude recovery, multi-task conditions, target recognition, and formation flight.

DEVELOPMENT OF THE PERFORMANCE BATTERY.

It will be noted that several of the above tasks are segments of actual flight requirements (ILS Landing, unusual attitude recovery, etc.). Others, on the other hand, are activities that are carried out across many different conditions of flight demand (e.g., monitoring under high workload, timing tasks, etc.). This suggested that the suite of tests that should be developed should contain two different kinds of task. One would attempt to simulate the flight task with some degree of fidelity, utilizing the elements of that task that were critical to success or failure. For instance, the ILS landing could be mimicked, and essential metrics which would ultimately reflect the success or failure of the landing could be obtained.

Since the second kind of task involved skills that were critical in several kinds of flight segments, it was clear that a generic test of those skills would have to be "synthesized". This means that the critical skills involved in those tasks would have to be incorporated into a test that might not look like any actual real world performance. Such synthetic tasks would have to validly tap one or more human resources directly related to the critical tasks defined above. Thus, "timing" would have to be measured with a test that, although it did not "look like" any real-world task, still captured the timing skill that is required in any real world task.

After analysis of the skills required by the critical tasks, a suite of ten task procedures was developed. The "simulation" tasks included: ILS approach, gunsight tracking, unusual attitude recovery, aircraft recognition, pitch/roll capture, and flight simulation maneuvers. The "synthetic tasks" included: motion inference, perception of relative motion, precision timing, and peripheral monitoring. Together, these test procedures constitute the Acceleration-Performance Assessment Simulation System (A-PASS). The A-PASS battery, when actually developed, will provide measures of virtually all of the skills required by the critical tasks listed above. This development should resolve the second of the three major questions phased in this program.

DEVELOPMENT OF OMI MEASURES.

The third question dealt with the need to convert laboratory metrics such as those generated by the A-PASS battery into actual measures of OMI. The techniques chosen to do this involved utilization of a new combination of computer simulation modeling techniques. A distinction was drawn between computer models that yield measures of human performance based on a network of interrelated performance activities, and those other types of models that simulate physical systems such as aircraft. The former models essentially take into account the trial-to-trial variability of human performance, and also the interrelated nature of complex performance strings. They output probabilistic sequences of performance times, accuracies, etc., for a large number of iterations of the complex behavior. The second type of model is more directed to determining the result or impact of a given activity or set of conditions on the physical system involved. For instance, a model of an F-16 might determine the position of the aircraft ten seconds after the pilot commanded a 6G turn (given a certain altitude, heading, airspeed, etc.).

It was hypothesized that if the performance values that resulted from testing with the A-PASS battery were entered into a network model of a given mission, the model could generate a large number of such missions. If data were collected with both non-stressed operators and operators who had been exposed to acceleration stress, two distributions of missions could be generated by the network model. These would provide realistic ranges of operator performance under the two experimental conditions. Of course, the output of the network model is still generic in the sense that it is not related to any specific system or aircraft. It simply represents a general description of how a non-degraded and possibly degraded set of pilots would perform a particular mission.

To determine the actual effect such degraded performance would have on a specific aircraft and mission, the output of the network model is used as input to a systems model. For instance, the question may be "What will be the effect of a given G profile on performance in the F-16 aircraft doing a pop-up ground attack mission?" To answer this, the output of a thousand iterations of the performance model could be used as input to an F-16 systems model performing that mission. The net result would be an estimate of the actual effect of the G profile on a representative population of pilots flying that mission in that aircraft. This would constitute a measure of OMI.

In Phase I, a simple demonstration of how these activities would be carried out was completed. Using hypothetical data, the MicroSaint network model was exercised to yield 37 iterations of a pop-up air-to-ground attack. The performance values obtained were then used as input to the BLUE MAX systems model of the F4-E aircraft. "Undegraded" pilots achieved a probable error (CEP) of 25.0 ft. (SD = 10.8), while the CEP for the "degraded" pilots was 50.1 ft (SD = 21.9).

The above approach results in a complete methodology for providing the military community with OMI measures for any mission or system for which a systems model exists. Since such models exist for virtually every system, the approach addresses the third general question raised above.

The basic design and demonstration of the A-PASS system carried out in Phase I sets the stage for the full development and production of the system. Although there are still problems to be solved before the final system can be used by researchers in the field, these all appear tractable within the context of a Phase II program.

1. INTRODUCTION.

Flight in modern fighter aircraft intrinsically results in the body being subjected to extreme physiological challenges that can have performance degrading and even life-threatening effects. Unusual and extreme vestibular stimulation, heat, and high-speed angular acceleration constitute only a few of these challenges. The field of aerospace medicine has long recognized the need to quantify and protect against adverse effects of such stresses on the pilot and on the operation of the aircraft.

As a result of this potential threat, a great deal of effort has been expended in centrifuges and other dynamic environments directed to ameliorating many of the physiological insults generated by high speed, high altitude flight. Although these studies have resulted in remarkable success in addressing the physical and physiological effects of such flight, there has been considerably less success in addressing more subtle issues of human cognitive performance limits in these environments (Von Gierke, McCloskey, and Albery, 1991: McCloskey, Tripp, Chelette, and Popper, 1992). In other words, there still is no precise way to quantify the potential performance decrement caused by physiological stress on the pilot, or even the degree to which performance may be degraded by the very techniques that are employed to protect the pilot from the physical threat.

The difficulties in measuring human performance in such unusual environments can be traced to the nature of the stressor environment itself (including the intrinsic variability of the pilot population), and to the complex nature of the prediction that must be made from a limited sample of behavior. As the stressor becomes more intense, with accompanying shortening of subject exposures, the researcher is forced to retreat to more and more basic (and indirect) measures of performance. This inevitably results in an increased use of laboratory measures that have a decreased meaning to the operational community. Although there have been creative attempts to bridge the gap between laboratory and "field", it is an unfortunate fact that no approach to performance measurement in high physiological-stress environments has provided data phrased in the operationally meaningful terms demanded by the system designer or field commander (see review by Perez, et al., 1987).

1.1 BACKGROUND.

The search for appropriate cognitive performance measures in the centrifuge has been long and tortuous. Beginning with a near loss-of-consciousness by a pilot in the Pulitzer Air Races about 1920 (Burton, 1986) there has been interest in the effect of increased $+G_z$ (head to foot) on human performance. This has resulted in a variety of testing procedures being used in centrifuge studies, and has produced a set of general conclusions about the overall performance effects of angular acceleration (especially $+G_z$ acceleration) on the human. Representative (selected) procedures and results are reviewed briefly below.

1.1.1 Physiological Indicators of Performance Capacity.

Early on, it was recognized that one reliable indicator of the pilot's ability to perform in the aircraft could be obtained by assessing the degree to which the individual was approaching loss of consciousness (LOC) under various stressor conditions (Burton, 1986). The underlying indisputable premise of this approach was that conscious processing of information is necessary for virtually every task in the aircraft. Therefore, LOC should be etiologically related to the person's performance level.

Techniques to monitor the physiological phenomenon of LOC, at least crudely, were relatively easy to define, if not to implement. Invasive blood pressure measures were used by early researchers, and these, plus non-invasive techniques to measure pressure and flow (e.g., Doppler techniques) continue to be used routinely (McCloskey et al., 1992). Oximeters located non-obtrusively in the pilot's helmet or face mask can reflect infrared and visible red light through arterial beds, and can give precise indications of blood oxygen saturation levels ($SaO2$)(Tripp and Albery, 1987). Since these measures are related to the degree of pressure in the hydrostatic column, and to the arterial oxygen saturation (McCloskey et al., 1992), they should reflect the pilot's capacity to perform the mission.

In fact, these 'indirect' physiological approaches may provide a more direct measure of the pilot's capacity to perform the mission than direct cardiac monitoring (EKG, echocardiography, or impedance cardiography), although one or more of these techniques are routinely available in centrifuge research. The same may be said of direct electrophysiological measures of the brain's activity, the EEG. Although such measures have been successfully made in the centrifuge (Berkhout, O'Donnell, and Leverett, 1973), and in aircraft (Sem-Jacobsen, 1961; Wilson, Purvis, Skelly, et al., 1987), the operational impact of these changes have not been assessed.

A more subjective, but still physiologically-based index of the pilot's level of consciousness (related primarily to blood movement in the hydrostatic column) can be obtained through measures of peripheral and central light loss (Fong, 1992). This is measured by having the pilot indicate when he or she can no longer see certain lights in a semi-circular bank of horizontal lights. In fact, some degree of standardization of the pilot's ability to see peripheral light was developed, and can serve as a performance monitoring technique (Coburn, 1970). Theoretically, loss of a certain degree of ability to see these lights provides a good physiologically-based end-point for assessing the individual's ability to perform in an aircraft.

The problem is that in the long-term, physiological indicators of performance capability, while well-grounded theoretically, do not perfectly correlate with performance prediction. This is true for a number of reasons. First, blood supply to the brain is only one factor influencing the pilot's ability to perform under G-loads. Mechanical effects on the body, such as neck strain, arm-muscle incapacitation, etc., are not directly measured by cardiovascular or brain physiology. Further, experienced pilots, as well as direct observation, reveal that certain conditions that precede severe LOC might be experienced during flight with minimal effects on actual performance (Sem-Jacobsen, 1961). In other words, pilots experience short periods of these conditions, but are still able to perform a mission successfully. Finally, "psychological" factors such as stress and anxiety or extreme motivation have been shown to affect centrifuge performance either negatively (Bennett, 1985) or positively (McCloskey, et al., 1992). It is clear from the statement above that, while physiological measures are indeed valuable, it would be desirable to have more direct performance-based measures.

1.1.2 Performance-Based Measures.

Researchers have endeavored to develop "synthetic tasks" that resemble, but are not identical to, the actual cockpit tasks. These are based on models of performance resources, and therefore could be said to require the same skills as the cockpit tasks. Performance on these synthetic tasks could therefore be theoretically generalized across many flight tasks and specific questions. Grether (1971) reviewed the early literature on synthetic task performance under $+G_z$ stress, and reported significant and generally progressive effects on visual functions (absolute and differential thresholds, visual acuity, and dial reading), reaction time, reaching and control activation, and tracking. Although only two studies reporting strict cognitive functions were reported, Grether proposed that cognitive function appeared to be more resistant to $+G_z$ stress than sensory or manual functions.

Perez (1986) provided an excellent review of selected but representative performance measures which have been used under sustained G_z acceleration. In general, it was noted that higher cognitive functions of maze solving and spatial visualization (the "Manikin Task") were not affected up to very high ($6\ +G_z$) levels. On the other hand, tracking behavior (which accounted for six of the eight studies reported) was significantly affected in most cases, although to varying degrees and at varying G-levels (see also Von Gierke et al., 1991). Chambers and Hitchcock (1963) showed that accelerations up to $5\ +G_z$ do not affect immediate memory, but that $7\ +G_z$ reduces memory performance. Earlier, Frankenhouser (1945) had also shown that longer-term exposures (2 to 10 minutes) to $3\ +G_z$ was necessary to produce significant effects on a multiplication test. Albery, Ward, and Gill (1985) demonstrated that performance on a maze task was not affected by forces from $1.5\ +G_z$ to $6\ +G_z$, although subjective estimates of workload increased.

The most frequently found effect of acceleration appears to be some form of tracking behavior. The logical operational assumption underlying utilization of tracking tasks is that they involve many of the skills involved in piloting behavior. Therefore, results from these tests can be used to infer real-world performance. This is particularly true if a well defined theoretical model of the tracking behavior is employed. For instance, the cross-over model (Jex, McDonnell, and Phatak, 1966) provides an opportunity to analyze the individual's tracking behavior into "elementary" components (e.g., describing functions). By relating these elements to the requirements of flight, it should be possible to determine what the real-world effects of a stressor would be by extrapolating the decrements seen in the tracking task to the real world requirement.

Unfortunately, this extrapolation from laboratory to real-world has not proven to be as simple as it first theoretically appeared. First, engineering models of closed loop control are essentially linear models. A problem arises when the tracking requirements of high speed flight are combined with multiple performance requirements not involving visual motor control. As the requirements of flight become more cognitive in nature, the validity of extrapolations from tracking studies to actual flight conditions becomes even more suspect. In other words, while closed-loop tracking behaviors can be measured with admirable precision, and while they may allow prediction to certain limiting aspects of flight performance under stress, they cannot be easily related to a real-world flight which involves many non-linear cognitive components, or used to predict the pilot's performance in a complex cognitive tasking situation.

In summary, it would appear from the above brief review that (using terminology more current in cognitive psychology) the functions that have been most often shown to be affected by acceleration have involved sensory input and motor output. Cognitive functions involving working memory have been tested less often, but with little significant effect.

Recently, there have been innovative attempts to incorporate other types of synthetic performance tasks into the centrifuge environment (see, for instance, the Sustained Acceleration Performance research Plan for AL/CFBS). State-of-the-art performance batteries, containing both multiple performance measures and physiological assessment, have been proposed for use on the centrifuge. These tasks provide extremely precise and meaningful data concerning the effect of a specific stressor on the individual, and certainly represent a quantum leap over past approaches. However, although these approaches will certainly provide an enriched data base, there is still no coherent approach suggested for extrapolating these laboratory results to the real world. In effect, <u>no generally accepted methodology currently exists for carrying out a coherent integration of laboratory results into an operationally meaning context.</u> Data are still reported in terms of experimental reaction times, percent correct, RMS error, changes in remnant function, millisecond changes in brain response, microvolt changes in skin conductance, or other significant but operationally intractable results. The operational commander is at a loss to interpret the practical meaning of these changes.

Simply, all of the above approaches fail to answer the basic questions found to be most important by the operational community: (1) "In actual flight, what is the probability that I can perform this particular acceleration maneuver and survive?", and (2) "In actual flight, what is the probability that I can perform this particular acceleration maneuver and successfully carry out the range of tasks which I may be assigned?"

The present effort was directed to answering these questions. First, meaningful synthetic performance tests which could be used in the centrifuge were defined. Second, a methodology for defensibly extrapolating those techniques to the real, operational world was demonstrated. In this way, it was shown that basic performance measures collected in laboratory settings could ultimately result in metrics directly useable by the operational community.

2. OVERVIEW OF THE APPROACH TAKEN

There were three major problems that had to be solved if performance assessments in these environments were to be related to operational requirements. Phase I of the present effort addressed these problems.

a. First, the range of operational missions to which data must be applicable is extremely wide and diverse. In short, there is simply no way that an experimental approach can test every possible operational condition of interest! Stated differently, although it is recognized that high-fidelity simulation can provide reasonably precise and meaningful data about a specific mission or field condition, this approach is prohibitively expensive and time consuming. It is literally impossible to create simulations for all of the situations of interest to the Air Force. A less expensive, more generic approach that still produces precise and meaningful data, is required.

The approach taken in Phase I to address this requirement essentially involved the creation of part-task simulations of "critical tasks or mission segments" that were selected by experienced pilots as representing crucial tests of the pilot's ability to survive and carry out the mission successfully. In other words, these critical tasks were those which the pilot needed to perform adequately to survive and prevail. Implicitly, of course, this assumes that if the pilot could carry out these tasks effectively, less demanding tasks could also be carried out with equal success. To the extent that these tasks were actually critical and representative, they drastically reduced the magnitude of the problem described above. One could now probe only these tasks, with some assurance that results on these probes would provide the "boundary conditions" upon which decisions about a vast array of operational situations could be made.

b. The second problem involved development of an actual set of laboratory measures. Once the critical tasks were defined, it was possible to analyze the basic skills they required. Although there is no universal agreement on the terminology or underlying theoretical model relating complex tasks to basic skills, the field of performance assessment has made significant progress in this regard during the last fifteen years (O'Donnell and Eggemeier, 1986). Fundamental breakthroughs have been made by the multiple resources model of Wickens (1980) and other information processing models (O'Donnell, 1993).

Using these approaches, as well as contributions of Shingledecker (1990; 1984), Englund, Reeves, Shingledecker, et al., (1987), O'Donnell (1991; 1993) and others in the area of performance testing, basic skills involved in real-world complex tasks can be defensibly identified, and a finite number of relatively orthogonal probes or tests of those skills can be defined. This was done in Phase I. A test battery of performance probes was conceptualized which targeted those specific skills the operational community had determined to be absolutely critical to fighter pilot survival and successful mission completion. The tests comprising this battery are described in detail later in this report.

It is important to recognize that this approach is relatively rare in the field of performance assessment. The concept of "critical tasks", if used at all, has typically been limited to generalizations (e.g., "piloting involves tracking ability, so we will measure tracking"). To our knowledge, no one has previously attempted to discover the boundary range of critical tasks, in specific terms, which would answer the question: "Can the pilot satisfactorily perform all the specific tasks which are necessary for survival and mission accomplishment?.

The successful solution of this difficult task, through the development of a test battery specifically relevant to critical piloting tasks, constitutes the first major innovation of the Phase I effort.

c. The third problem involved the actual translation of experimental laboratory results into terms that would be practical, useful, and meaningful to the operational Air Force community. Performance researchers have produced a formidable array of test procedures which are capable of probing human skills with extreme precision and reliability (Shingledecker, 1984; Englund, Reeves, et al., 1987; Perez, Masline, Ramsey, and Urban, 1987; and others such as the NATO STRESS battery). The data from these existing test batteries provide excellent assessments of the person's present capability in any number of well-defined skill areas. Obviously, then, there is no lack of precision in our ability to measure skills, once they are defined as in the above paragraphs.

In point of fact, however, there are problems with the extrapolation of the results of tests such as those noted above to the "real-world". Necessarily, most performance tests result in data being expressed as either speed of response ("reaction time"), or accuracy ("percent correct"). In relatively rare instances, more sophisticated metrics can be derived from the performance results (e.g., describing functions for tracking behavior, or interval measures of accuracy such as "degrees of error".) Again, it is noted that such metrics provide the most appropriate and precise description of the behavior being tested. However, they result in measurement units (e.g., reaction time in milliseconds; percent correct versus incorrect; lead, lag, remnant terms in an equation; magnitude of error; etc.) which are, at best, meaningful only in the experimental situation.

While scientifically (and statistically) interesting, these metrics fail to address the meaning of the data to the operational community. If a statistically significant difference in reaction time of 100 milliseconds is shown in a pilot after a given stressor, the military commander may still validly ask: "So what!", or "How will this affect my ability to carry out the mission?"

In view of this, an explicit approach had to be developed which would allow rational and defensible extrapolation from experiments that produced scientifically esoteric data, (frequently involving a small number of experiments, with limited numbers of subjects and exposure durations) to the mission and pilot populations of interest. To achieve this, the approach had to not only specify the extrapolation procedure objectively (i.e., mathematically), but also had to result in data that were phrased in terms easily understood by the system designer or field commander.

The Phase I approach to this problem was to utilize a combination of existing and newly created computer modeling techniques. These models are simulations that utilize mathematical descriptions of system dynamics (e.g., the actual response of an aircraft or missile to a change in control input), or of human behavior (the inter-dependent effects of one behavior on a series of later behaviors). To our knowledge, these two types of models have never before been combined to address the operational questions of interest to the high-stress performance community. Models of aircraft systems have been developed extensively by the Government, while models of human performance have been developed and marketed by commercial sources, although frequently under Government contract. The Phase I results described below demonstrated that, by combining the capabilities of models of the human with models of the actual aircraft system, the output would defensibly provide probability estimates of the operational outcome of any defined mission. In this way, the end-user of research would not have to extrapolate from esoteric laboratory data (such as reaction time or percent correct), but would have data that were immediately useable and meaningful.

3. OVERALL CONCEPT OF THE A-PASS PACKAGE.

As noted above, the need for an integrated laboratory/modeling performance assessment system arose from the fact that no current system possesses both the sensitivity and operationally meaningful metrics demanded by the military community. The Acceleration-Performance Assessment Simulation System (A-PASS) is proposed as the solution to this problem. In order to fully understand why this is so, it is desirable to discuss the underlying rationale and assumptions of the overall approach. These are presented below in the context of specific performance assessment problems.

3.1 THE ISOLATION OF SKILLS IN A COMPLEX PERFORMANCE.

The fundamental assumption underlying A-PASS is that the skills required in a complex performance situation can be individually isolated and studied. In other words, no matter how complex the task required of the human operator, its performance essentially depends on the operator's skill in <u>one or more areas or domains</u>. It is the premise of A-PASS that, by measuring the operator's performance in each of these relevant domains, one can arrive at a defensible prediction of his or her performance <u>in the complex behavior</u>. More precisely, one can determine whether the operator's performance in the complex behavior will be degraded or decremented, and arrive at a defensible prediction of the degree of decrement.

A simplified but operationally meaningful example is shown in figure 1, which illustrates some of the skills involved in a "missile defense" maneuver. This is required by the pilot when attacked by an air to air missile. On the surface, it might appear that this is a complex series of behaviors that could only be studied with high fidelity simulation. However, if one analyzes the elements of the missile defense maneuver, it is possible to see how the essence of the complex task can be extracted and studied.

The first task of the operator in this maneuver is to detect and respond to the threat rapidly and appropriately. Typically, the detection task is performed either by a sensor on the aircraft or by another friendly aircraft in the vicinity. It therefore does not require a high level of sensory acuity. However, when a warning tone or voice communication occurs, the individual must respond as quickly as possible with an appropriate breaking maneuver and activate appropriate systems of countermeasure (number 1 in figure 1). This is essentially a choice reaction time task to a discriminative stimulus that is well above threshold. If one probes the pilot's performance with a generic or synthetic choice reaction time task, under various stressor conditions, it will be possible to state whether any of the stresses caused a decrement in the pilot's reaction time capability to this type of situation. Further, since such a decrement is probably physiologically caused, the quantification of the decrement should be generalizable to many other real-world situations requiring the same type of response.

However, in the missile defense maneuver, it is not sufficient for the pilot simply to react with the appropriate initial maneuver. It is then desirable to perform a

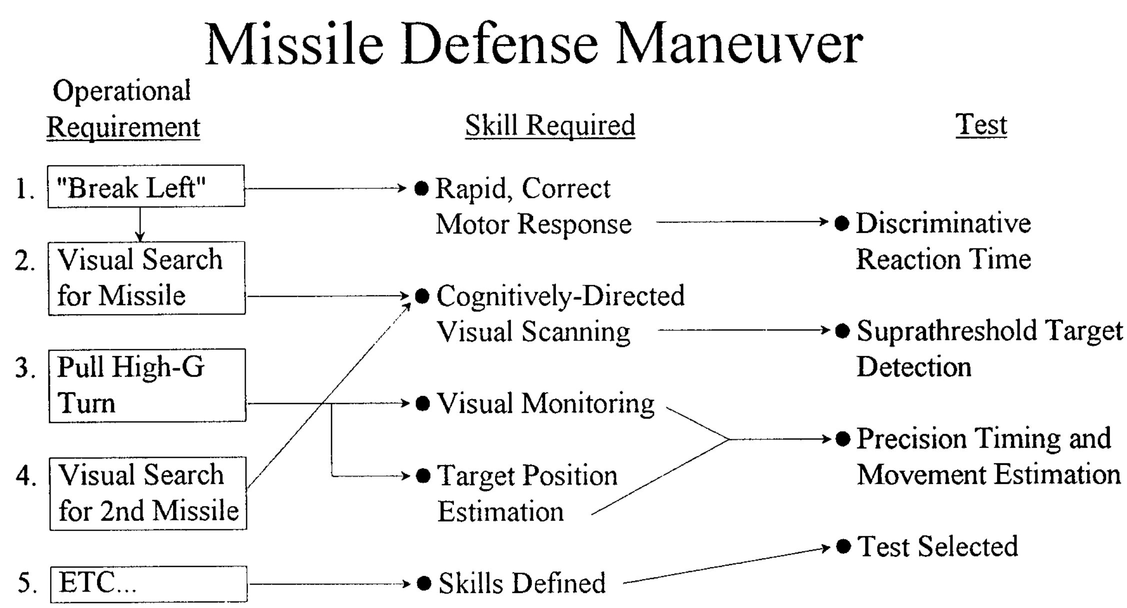

FIGURE 1. SCHEMATIC REPRESENTATION OF SKILLS INVOLVED IN THE "MISSILE DEFENSE" MANEUVER

visual search for the missile itself (number 2 in figure 1). Again, visual search of this type is a task which can be simulated generically under a laboratory condition through use of a target detection task in which the subject has some prior knowledge of the probable location of the target.

Once the missile is visually located, the pilot's task is to perform an appropriate maneuver that will cause the missile to miss the aircraft (number 3 in figure 1). This requires a precision visual motor/timing skill in which, often enough, the pilot loses sight of the missile. He or she still must estimate when the missile has passed because, as soon as possible, the pilot wishes to search for a second possible missile that might have been fired (number 4 in figure 1). This task may or may not be the same as that shown in number 2. If so, the same test may be used to assess the skill. Otherwise a test could be found that probes the appropriate skills. This type of analysis continues until all of the critical performance requirements of the missile defense maneuver have been identified (number 5 in figure 1).

The sequence of individual behaviors that make up the entire missile defense maneuver, therefore, can be seen to consist of a number of individual elements, each of which requires definable skills on the part of the pilot. Each element in that sequence of behaviors can be defined and studied under laboratory conditions, and the entire sequence of real world behaviors can then be described in terms of a sequence of required skills. This is a basic foundation upon which the A-PASS concept is based.

3.2 ACCOUNTING FOR INTERDEPENDENCIES AND VARIABILITY IN PERFORMANCE.

Although the identification of skills, and the development of laboratory probes to assess those skills constitutes one major part of the specific A-PASS development, this alone would not provide a great deal more than could already be done with existing technology. Several other assumptions and technical developments are required.

First, it should be evident that in the real world, individual components of a complex task do not operate in isolation. There are significant interdependencies among the elements. For instance, if an individual is late or inaccurate early in the event sequence, greater accuracy will be required later in the sequence in order to compensate for the early error. This leads to a second major assumption of the A-PASS system. It is assumed that such interdependencies can be appropriately described and modeled by probabilistic and statistical methods.

Specifically, a class of computer models exists that can be used to account for interdependencies among tasks in a complex performance. In such models, each behavior or performance is dependent on the state of previous performances. In other words, as noted above, these models treat complex performance as a series of sequential actions or specific performance requirements. This series constitutes a "network" (and for this reason, this type of human performance model will be referred to as "network" models, although they have been identified by different terms in the literature). Each required performance in a complex task such as that shown in figure 1 constitutes a "node" in the network. On a single "run" the individual "performs" each node, in the sense that the computer selects quantities (time, accuracy, etc.) that describe how a human responds to the task demands. These computer selections are made on some probabilistic basis. For example, the distribution of "normal" reaction times for a given task is entered into the computer, and on a given "run" the computer selects a reaction time -- most often selecting one that occurs most frequently in the distribution.

However, the nodes in a network are not isolated random performance selections by the computer. Nodes can be linked together by mathematical statements that describe their interdependency. For example, one node may not be activated until the simultaneous activity required by two previous nodes is completed. In another case, a node may be assigned a different distribution of performance probabilities based on the fact that a previous node was performed slowly (or poorly). In other words, the interdependencies of various performance requirements in a complex task are reflected in a network model.

In the A-PASS system, the use of network models in conjunction with the kinds of performance data that will be generated by the A-PASS test battery will permit conversion of basic laboratory performance data into estimates of overall performance in complex scenarios.

A second benefit derived from use of network models is based on the fact that they account for variability in human performance. Humans, of course, do not perform exactly the same every time they do a task -- there is a distribution of performance around some measure of central tendency. This is relatively easy to describe if we are dealing with a single task. We simply determine the shape of the distribution (e.g., normal, poisson, rectangular, etc.) and then use appropriate measures of central tendency and variability. However, when a complex task involves many inter-connected tasks, the job becomes more difficult. There can be no "single" set of numbers that accurately captures the fact that the person may have performed each of the tasks a little differently every time it was done. In point of fact, under such conditions, one is forced to use a large number of "runs" (instances of the person performing the complex task) and then compute probability values based on that large performance data base.

Of course, it would be prohibitive to do this many runs experimentally. Network models, however, can be iterated thousands of times in a very brief period, sometimes in a matter of minutes. On each iteration, the computer "selects" a performance time or accuracy for each node, based on the pre-determined distribution of performances (e.g., from tests such as A-PASS). Thus, for a single "run", a set of performance values is probabilistically determined (limited by the interdependency statements described above). Most of the time, of course, these values will be near to the central tendency of each specific element in the complex task. Overall, the distribution of performance obtained over many runs will reflect that that would be seen in variable humans. Thus, the network model has taken human variability into account.

In the A-PASS system, such network models are used to generate a large number of specific sequences of events, each of which is possible. For instance, if such a network was developed for the missile defense maneuver, a single "run" through the network would describe the complex performance pattern of a pilot performing a single missile defense maneuver. A thousand "runs" through the model would give a distribution of performance simulating a thousand flights by pilots of defined performance capability.

3.2.1 Conversion of Laboratory Data to Operational Military Impact (OMI) Measures.

Network-type models have been in existence for at least 25 years, and have achieved a high degree of sophistication. The problem, however, is that they are usually generic. They describe the outcome of a series of events simply in terms of some generalized metric -- such as time to complete each event, or overall accuracy. To be useful to the operational community, such generic outcomes must be translated into real-world terminology. This means that a change in reaction time, or a particular amount of error in a sequence of events must be converted into terms that describe the effect of such time or accuracy errors on a real world system.

An example might clarify this need. It is not enough to know, for instance, that a pilot might be degraded in reaction time by 100 milliseconds. The real question of interest is what is the effect of that 100 millisecond difference on the system and/or mission that the pilot is performing? A 100 millisecond difference in response time in most civil transport aircraft missions would probably not have an operationally meaningful impact. However, a 100 millisecond difference in a high-speed, high-performance fighter aircraft could have a significant or even catastrophic effect in some situations.

The next task in the A-PASS approach, therefore, is to convert the generic metrics revealed by the network models into operational military impact (OMI) statements. To do this, it is necessary to have appropriate descriptions of the system of interest. If one is attempting to predict the OMI of pilot decrement on a specific aircraft, it is obviously necessary to have an aeronautical model of how that aircraft flies. In this way, a 100 millisecond difference in reaction time can be converted to a difference in the aircraft state vector. A delay of a 100 milliseconds in an F-16 may not result in the same position of the aircraft as the same delay in an F-4.

In the A-PASS system, this translation is accomplished through the use of a variety of "systems" models (again, as with the term "network" models, the term "systems" models is used rather loosely to refer to computer models of the aeronautical dynamics of various actual aircraft or missile systems). Typically, they require input regarding the control actions or aircraft positions at critical points in the flight profile. Then, using the actual aircraft or missile response characteristics, the model calculates where the aircraft will be (and where it will be pointed) at X amount of time after that point. This introduces system-specific modeling, in the sense that the same control input (at exactly the same time) in an A-10 or an F-111 aircraft will result in different effects on aircraft positions 5 seconds later.

Systems models are extensively available for virtually every military system that might be of interest. In fact, the Survivability and Vulnerability Information Analysis Center (SURVIAC), located at Wright-Patterson AFB, is the repository for many of the most important systems models.

The key point in the A-PASS system is that control inputs (or positional commands) generated by the network models can be used as input to systems models to determine what a specific real-world system might actually do.

The output of the network models described above consists of a large number of specific instances (e.g., individual flights, or missions, or mission segments). These are described in terms of the speed and accuracy with that the pilot commands the system at particular performance nodes. The systems model takes these control commands and determines what an aircraft or other system might actually do in space, given those control inputs. In other words, the systems model takes the individual flight, in terms of speed and accuracy measures selected by the network model, and determines what a specific real aircraft would do in response to those commands. For that particular instance, therefore, the system model reveals where the aircraft ended up as a result of all the sequences of inputs it has received. This state description of the system can then be easily translated into a mission outcome statement, or OMI. For instance, a separate model can utilize the state description to determine where a bomb would be delivered, or whether a missile would have hit the aircraft, or whether a given maneuver of the aircraft would have resulted in a crash, etc..

This, then, represents the final output of the A-PASS system. In summary, reasonably small samples of empirically derived values of pilot performance in critical skills are used to generate a large number of examples of realistic combinations of events in a complex behavior. These examples are then converted into actual systems effects, and probability distributions of OMI are generated. If the basic performance data are gathered on individuals in a non-degraded condition, and then again when the individual is degraded by some stressor, the A-PASS output will allow the operational commander to determine empirically what the operational effect of that stressor might be on any mission.

4. RESULTS OF PHASE I EFFORTS.

4.1 OVERVIEW.

The overall Phase I effort to achieve the ambitious requirements of the approach described above involved several integrated steps. First, it was necessary to determine what kinds of operational missions the pilot community considered "critical" in the sense described earlier (i.e., those mission segments that constituted "boundary" conditions -- those that a pilot had to be capable of performing in order to survive and prevail). This was done in a meeting in which pilot consultants, as well as Government personnel, developed a series of mission descriptions. Next, it was necessary to determine the kinds of skills that were involved in those missions, and to develop a battery of tests that probed those skills. This was done by NTI scientists, and was verified in another meeting in which pilot consultants and active Air Force pilots evaluated the proposed tests and made suggestions regarding their face and content validity. Finally, an intensive modeling effort was carried out to select appropriate network and systems models, and to provide a prototype demonstration of how such models might be employed in a final A-PASS system to provide OMI metrics to the operational community. Each of these efforts is documented in detail below.

4.2 DETERMINATION OF OPERATIONAL "BOUNDARY" MISSIONS.

As described earlier in this report, "boundary missions" are those that experienced pilots' feel represent operationally meaningful tasks that are critical for mission survivability and success. The need to define such tasks is based on the reality that not all mission tasks can realistically be probed by any predictive performance system. Therefore, the concept introduced by A-PASS is to select specific missions or mission segments that are general enough that, if the pilot could perform them, he or she could certainly perform many other tasks. In this sense, "boundary conditions" were conceptualized as prototypical critical tasks.

Under the constraints of this definition, the two experienced pilot consultants on this project suggested a number of missions and mission segments that they considered to be prototypical. These were based not only on their own experience, but also on a review of the recommendations developed from a workshop at Wright-Patterson AFB (AL/CFBS) in that experienced, active Air Force pilots identified pilot tasks that they considered to be critical. This latter list of critical tasks was all-inclusive, in the sense that it included all of the tasks that the pilots considered important to flight. In the present effort, an attempt was made to eliminate redundancy in these tasks, and to develop a list of "critical" missions that were relatively independent of each other. Thus, the original list suggested at the AL/CFBS workshop was reduced to nine mission types that were both critical and somewhat independent in terms of skills required. A synthesis of the results of this analytical effort is summarized in Table 1, and are discussed below. Obviously, no two missions are totally orthogonal, and the terms used in the Air Force workshop are different from those used for the A-PASS. Therefore there will always be some overlap among the missions selected as well as not having a straightforward identity between the A-PASS missions and the terms used in the workshop. However, we believe that this list of A-PASS "critical" missions constitutes a basic starting point of defining the boundary conditions, encompassing those discussed at the Air Force workshop, that must be evaluated if one is to predict to a wide range of operational conditions.

TABLE 1. COMPARISON OF CRITICAL MISSIONS FROM AL/CFBS WORKSHOP AND A-PASS EFFORT

SKILLS/TASKS IDENTIFIED AS CRITICAL DURING AF WORKSHOP

- Observe and kill bandit
  - Maintain sight
  - Maintain advantage
    - Manage energy
    - Achieve shot parameters
- Verbal
  - Communicate
- Motor skills
  - Initial missile avoidance
  - Detect second missile
- Visual
  - Acquire target
  - Recognize threat
  - Evaluate threat
    - Radar lock
    - Missile parameters
    - Bandit range
    - Etc.
- Awareness
  - Analyze situation
  - Check gas, airspeed, floor, etc.
- Cognitive
  - Develop plan (engage or leave)
  - Analyze bandit's maneuver

CRITICAL "BOUNDARY" CONDITIONS IDENTIFIED FOR A-PASS

1. Monitoring tasks
2. Collision avoidance
3. Missile avoidance
4. General timing ability
5. Landing
6. Unusual attitude recovery
7. Multi-task conditions
8. Target recognition
9. Formation flight

The specific critical tasks identified by the pilot consultants in the A-PASS effort are listed below and briefly described.

a. Monitoring Tasks: The ability of the pilot to monitor such things as altitude and attitude (both head up and head down conditions) was considered absolutely basic to survival.

b. Collision Avoidance: The ability to monitor the relative motion of an aircraft that is moving in relation to your own aircraft was considered critical. This closure rate was considered important, not only in head-on encounters, but also relative to the position of each aircraft in formation. For instance, during a high G turn, the movement of the friendly aircraft toward you is a critical performance parameter. Significantly, both pilot consultants agreed that this was important in both night and day operations.

c. Missile Avoidance Tasks: This was considered to be a complex critical task involving several specific skills. Initial detection of the missile involves both sensitivity and situation awareness. Once a missile is detected, it was considered important to measure the ability to keep sight of the target, even at high G loads. The actual breaking reaction time after a missile has been detected was considered critical. Finally, a tracking task, that involves controlling the aircraft's maneuver plane in relation to the missile constitutes an essential element of this overall task.

d. General Timing Tasks: The consultants discussed a large number of tasks that involve critical timing elements. Specifically, time delay in radar lock, time delay in missile launch, and time to pull an ejection handle were mentioned as critical survival and mission accomplishment elements. The pervasiveness of timing in the discussions suggests that this element of performance assessment should be given high priority. Specifically, differences in reaction time that are related to critical differences in the type of tasks demanded by the pilot should be extensively explored in this program.

e. Landing: There was a considerable amount of discussion and agreement that ILS landing constitutes a major critical element, and that the landing task provides a rich source of performance measures. Obviously, tracking behavior is clearly represented in an ILS situation. In addition, timing functions such as the flare path decision is important. In fact, several decision points during the ILS approach (abort, etc.) provide potential sources of performance assessment that could have general applicability in other flight segments.

f. Unusual Attitude Recovery: This represents a survivable-critical element that could be simulated reasonably well in the centrifuge. Clearly, it encompasses situation awareness, and includes many aspects of manual control and visual-motor coordination that could be generalized to other flight conditions.

g. Multi-task Conditions: It was pointed out that critical situations seldom occur in isolation. Therefore, the desirability of obtaining some measures of the impact of one task on another through a multi-tasking situation was discussed. It was recognized that this type of measurement approach would be somewhat complex and would require an innovative methodology. However, it was recommended that this multi-tasking approach be given some priority in the A-PASS development.

h. Target Recognition: The general question of the pilot's ability to discriminate targets from friendlies, or to identify friendly forces specifically, was considered important. Obviously, the timing and accuracy and such recognition for both were considered critical. Some attention was given to the way in that data could be gathered in the most generic form. Possible concepts ranged from presenting silhouettes, to a synthetic task involving patterns of light that the individual would have to identify.

i. Formation Operations: The general formation operation involves a wide range of skills that are considered to be flight critical. These involve such things as vectoring, three-dimensional tracking, decision making, collision avoidance, and monitoring. Again, while many of these could be tested in other ways, it was suggested that formation operations may represent an economical way to collect a large quantity of data that could then be related to other mission aspects.

4.3 DEVELOPMENT OF THE A-PASS TEST BATTERY.

The A-PASS test battery described in this section represents the result of several analytical processes carried out by NTI scientists. It is important to recognize that the process employed in this development is one that is generally recognized as desirable in theory, but that is not always carried out in practice. The development actually started out with an intensive analysis of the operationally desirable <u>end-point of the testing</u>, and then proceeded to develop tests that would contribute to that end-point. Too often, the testing field starts out by considering <u>that tests are available</u>, and then attempts to interpret those tests in terms of the operational end-point.

The process followed in the present case began with an analysis of the boundary conditions that had been defined in the first stage of the effort, and that were described above. Using current performance theory, the mission requirements described in Table 1 were analyzed into basic skill requirements. Obviously, this was not a simple exercise, since the complex requirements of the missions described in Table 1 were not always able to be decomposed into unitary skills. However, it was possible to conceptualize a finite set of skills (not always simple) that encompassed virtually all of the tasks defined by the pilot consultants. Somewhat surprisingly, this was a challenging but not excessively difficult task.

The next task was to develop a set of performance probes that defensibly probed the skills identified. The prime criterion for selection of these tasks was that they could be defended, on the basis of currently available literature, as measures of the skills required. This requirement dictated that the tests chosen be similar in all essential respects to well-studied laboratory tests. However, a requirement of nearly equal importance was that the tests have some degree of face validity. In other words, the tests had to resemble the operational tasks to that they would ultimately be applied. This dictated that the tests selected for the A-PASS battery be "new" in the sense that they did not simply use the stimulus material and response manipulanda of academic/laboratory tests, but that they contained the same "construct" validity of those tests.

The results of these analytical efforts by NTI are presented in detail below, in the form of detailed descriptions of each of the tests selected for the A-PASS battery. It should be emphasized, of course, that the specifics of the tests described here are purely analytical, as dictated by the Phase I nature of the effort. It would be hoped that additional experimental data would provide the basis for considerable revision of these analyses.

4.3.1 Test 1. Perception of Relative Motion.

4.3.1.1 General Description of the Procedure.

One of the more important skills required of the pilot is to be generally aware of the relative positions of one or more aircraft with respect to his or her "own ship". This is not a static skill requirement, but exists in a constantly changing geometry. The pilot must not only be aware of the instantaneous relationship between outside aircraft and own ship, but must be performing manual control tasks as a result of that relationship. The pilot's goal or final desired relationship also determines how to carry out this maneuvering. Combat maneuvering against one or more targets represents the quintessential example of this set of skills.

It is therefore desirable to probe the pilot's ability to display these skills. In a high fidelity simulation environment, it might be possible to set up an exact scenario that could be used for this purpose. However, this becomes problematical when one considers the fact that there would be a variety of optimal ways for the individual to carry out any given combat scenario. In addition, the task of creating a "smart adversary" in such a situation would make this type of approach extremely expensive and complex. Analyses would be similarly difficult.

Another type of task, however, could make the same skill demands on the pilot, while constraining his or her options. The operational task of joining up with a formation or another ship requires the same types of manual skills as any other air-to-air engagement. The difference is that the goal is clearly defined or definable (i.e., joining the formation as quickly and efficiently as possible) and the maneuvering options are likewise able to be mathematically described. In addition, the maneuvering of the other ship can be pre-described or determined, thus eliminating the need for artificial intelligence to be built into the system.

For these reasons, a formation join-up task was chosen for inclusion in the A-PASS test battery. Generally, the task described below, illustrated in figure 2, involves having the subject "fly" a target to another moving target. The subject will use throttle and stick inputs to control one of the targets in three dimensions (up-down, right-left, or a speed dimension). The subject's task will be to "join up" with the moving target as quickly as possible.

The major parameters that will be manipulated in this task are the initial starting positions of the two targets, and the maneuvering of the moving target. The subject will first see both targets on the screen in one of eight pre-defined

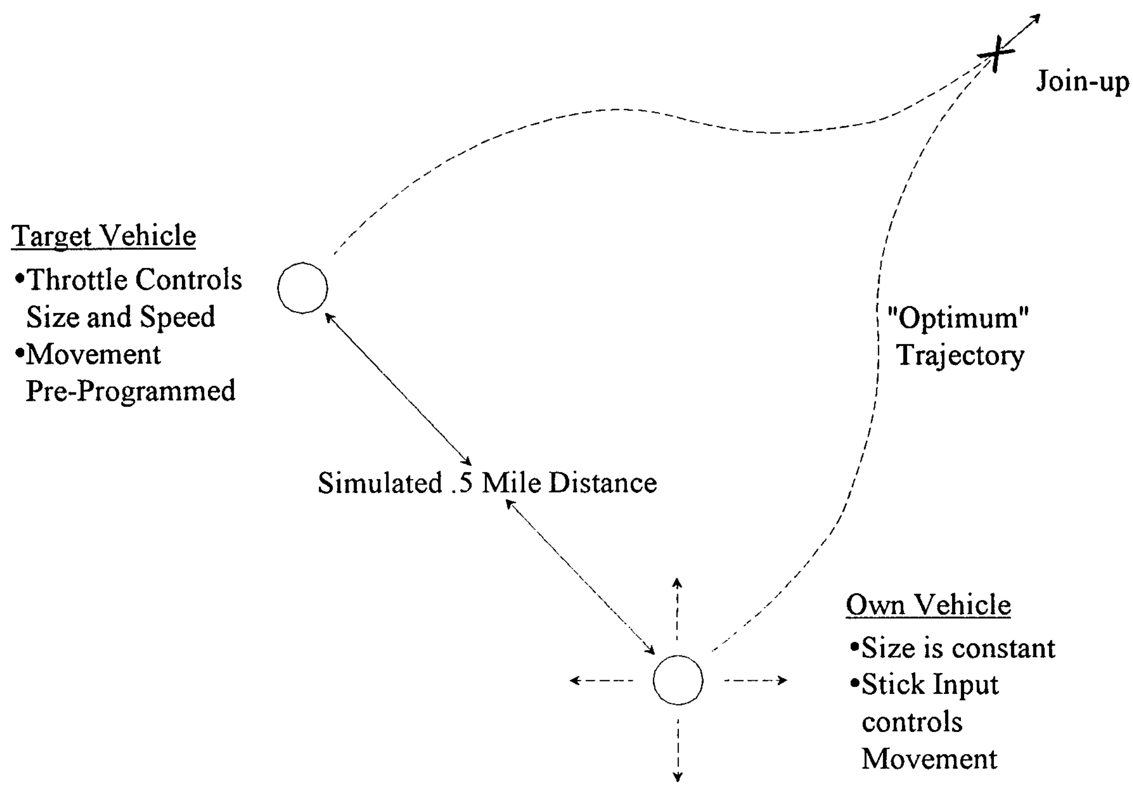

FIGURE 2. PERCEPTION OF RELATIVE MOTION TEST

starting positions. The goal will always be to make the two targets just touch each other (i.e., the goal is to join the two targets as quickly as possible without "crashing"). The moving target will describe one of eight trajectories, ranging from a straight line in the horizontal dimension to a rapid "jinking" maneuver. Timing will commence with the onset of the stimulus materials, and will end when the two targets touch each other. If the targets approach each other too rapidly, they will "crash", and this will also serve as a data point.

4.3.1.2 Construct Analysis of this Task.

The task described here is unique in its specific aspects. It cannot therefore be directly related to any known or previously studied skill tests. However, a construct analysis of the requirements of the task reveal similarities to a number of other tests of skill that have been reasonably well studied, and that form the basis for the underlying belief that the present task will provide an adequate and appropriate measure of the basic underlying skills.

Obviously, at its core, this task is a form of tracking. Although it is neither a pure compensatory nor pursuit tracking task, it shares elements of both of these traditional types of tracking. For instance, it would be expected that if a describing function could be generated relating the stimulus input to the subject's response, it would be found that lead and lag functions could be isolated. This would suggest that the subject treats at least some aspects of the task as a compensatory task.

In fact, adding the third dimension of speed introduces anticipatory requirements suggestive of second-order dynamics. In other words, the subject must anticipate speed changes as a function of the target's acceleration, rather than its instantaneous response or position. Typically, this drastically increases the need for lead in the subject's responses.

To the extent that scenarios require relatively smooth movement from the moving target, and to the extent that is predictable, the subject may treat this as a pursuit tracking task. Typically, this type of task requires smoother, more predictable responses. The reason for including this complex type of tracking requirement in the present case is that the real world task that is simulated here requires both of these types of behavior from the pilot. Decrements, or deltas, in performance on this task may be due to degradation of either of these skills of the subject (if in fact they truly are different skills). It is also argued that if the overall performance on this task is degraded from degradation in any of its underlying skills, the data can be legitimately extrapolated to any real-world tasks involving <u>any</u> of the underlying skills.

The present task also obviously contains elements of decision making and response speed. Since there will be 64 permutations of the possible scenarios that the individual will face, he or she will have to respond to the position of the moving target rapidly as it might change directions. The speed at that the person responds to both the initial scenario conditions, and to changes in the direction of the moving target will obviously determine, along with the accuracy of tracking, the total time to perform the formation link-up. Therefore, the task will be sensitive to changes in the person's dynamic decision making capabilities. Although this is a poorly studied skill, it is an extremely critical one in the flight environment, and relates to many of the flight skills noted in the Air Force workshop. It perhaps contains elements of what has been called attention allocation, or even schema development. As such, it would be expected that data from this task can be extrapolated to tasks that require continuous attention to an environment that could change at any moment. Therefore, it certainly would relate to formation flying and to combat situations.

Finally, the requirement that the individual make distance judgments based on size of the moving target suggests that certain aspects of visual perception might be probed. Specifically, what has been called "looming" detectors have been postulated in the retina. These permit the individual to respond relatively automatically to visual targets that are approaching. It could be hypothesized that any form of oxygen deprivation to the retina could have an effect on such detectors, and could affect visual perception in subtle ways. Thus, although the subject's task here is rather different from those in that looming detectors are usually probed, results of this task might be extrapolated to flying tasks that require sophisticated levels of visual perception. Combined with other tasks in the A-PASS system, it will provide sensitive data on the pilot's visual capabilities after or during exposure to high G's.

4.3.1.3 Detailed Specifications for the Task.

This is envisioned as a task lasting at least 30 seconds, up to as long as the experimenter wishes. Individual trials on the task will average between five and ten seconds. The number of trials required for stability will, of course, be determined during Phase II of this program. However, it is predicted that with adequate training, as few as three trials might yield stable results. Certainly, it would be expected that a stable result in a moderately well trained individual could be obtained in one minute of testing, encompassing ten to twelve trials.

The stimulus elements on the screen will consist of two circles representing aircraft (see figure 2). One circle ("own vehicle") will remain stable in size (approximately .75 inches), and stationary. However, control movements by the subject will cause the "other ship" to appear to change its relative position, using dynamics that approximate the response characteristics of an F-16 aircraft. The initial distance between this icon and the second one will simulate approximately .5 mile, or whatever distance would permit a potential link up in approximately five to ten seconds, assuming a terminal closure speed of about 10 kts. at 1000 feet.

The second circle will also be representative of an aircraft ("other ship", or target). This target will have been preprogrammed to move on its own. This movement will generally be relatively slow, to approximate the expected visual motion of a friendly ship that is waiting to link up with a formation. In most cases, the trajectory of this motion will either be a straight line or a fairly gentle curve. Periodically, however, this aircraft may make a sudden move up or down, with a change in apparent speed. This situation should model the situation in that the aircraft was required to make an evasive maneuver, or simply committed an error. In addition to this pre-programmed movement, the target circle will appear to move in response to the subject's own control inputs. In other words, the simulation will appear as if the pilot were observing the target circle from inside his or her own cockpit.

The size of this second icon will be controlled by the subject's throttle. Thus, the interaction between the joystick and throttle will determine an apparent "closure rate" between the two icons. As the subject flies toward the target aircraft, it will become larger (and appear to move faster) in proportion to how close the two objects are. This throttle control, therefore, obviously simulates the speed of one's own ship in approaching the other target.

The subject's task is simply to "fly" his or her own ship to the point where the wing tip is a predetermined distance from the target aircraft's wing. This is to be done as rapidly as possible without, of course, crashing into the other aircraft. The task will therefore require the subject to rapidly choose an optimal trajectory that will place the aircraft together. Once this determination is made and an initial flight path is initiated, the subject must determine the appropriate speed at each point along the flight path. Vigilance must be maintained, since the subject will be aware that the target aircraft occasionally makes unpredictable moves. If such a move should occur, the subject would need to recalculate a new trajectory and speed to minimize the link up time. (It is, of course, possible to introduce additional requirements or decision tasks. For instance, if the target aircraft makes a sudden jinking maneuver, the subject might be required to abandon the link up, and make a similar jinking maneuver.)

The precise definition of an "optimum trajectory" will be determined during Phase II, in consultation with pilot consultants and Government personnel. Nominally, it would be anticipated that the subject would be instructed to aim for a point behind the target aircraft, make an optimal turn, and approach it from the rear. However, decisions concerning this involve a number of factors that can only be determined after some prototype demonstration of the test is available.

The principal scoring dimensions of this task will be the time taken to link up, and the link up accuracy. At this point, we do not anticipate calculating such things as RMS error from the "optimal" trajectory. We also do not anticipate collecting measures of such things as simulated fuel consumption. These, however, could be considered as potential metrics from this task.

4.3.1.4 Potential Problems with this Task.

This task has been considered by our pilot consultants to have a great deal of face validity. They believe that it will yield appropriate measurement of skills that are critical in a wide variety of flight tasks. However, it should be noted that, as with any complex task that must be administered in a limited period of time, the amount of data that will be collected during one test is severely limited. In addition, this task obviously contains multiple skill elements, and the data will not reveal that elements are contributing to an observed decrement. Therefore, the test is obviously not intended to be completely diagnostic.

A related consideration deals with the type of scenarios selected, and their mix in any given test. Some scenarios (trials) will obviously be more difficult than others. Since there will be limited opportunities to present a large number of trials, severe imbalances in the difficulty level of the test from one administration to another could occur. This is not an unusual problem, but is exacerbated in the present case by the desire to use the data in computer models.

Several possible solutions to this problem will be considered during Phase II. First, the scenarios themselves could be equated in difficulty, so that any combination of them will be equally difficult. This may not be a desirable solution since it might limit the highest level of difficulty that could be introduced into the scenarios. A more reasonable solution would be to simply quantify the difficulty levels on at least an ordinal scale. Then, either automatically within the computer or through manual intervention of the experimenter, tests could be constructed with levels of difficulty that equated the overall difficulty of the test. In any case, these issues will be addressed during Phase II, and it would be anticipated that the final manual on the A-PASS system would provide definitive recommendations concerning how a test should be constructed from the various scenarios.

4.3.2 Task 2. Precision Timing Task.

4.3.2.1 General Description of the Procedure.

The pilot consultants on this effort identified several types of piloting demands that involve precision timing (general timing ability), especially visually directed precision timing. Decisions on weapons release in both air-to-air and air-to-ground situations, flare decisions, formation flight, and decisions to abort landings or other activities might fall into this category. The essential skill demand is that the individual visually monitor a changing situation, and decide at some critically identified point to initiate a motor action. A typical instance of this type of performance requirement could be when an enemy aircraft passes directly in front of your guns, and you need to decide exactly when to fire.

To probe this type of skill, the A-PASS battery will contain a task in that the individual must monitor what appears to be a rapidly moving target for a brief period of time. At some point in the path of the target's motion, an indicator will be inserted. This indicator is the point at that the subject should press a button on the stick. The distance and/or time error of the subject in precisely identifying the point at that the target should have been stopped will constitute the basic measurement parameter of this task.

The subject will see a curvilinear pattern of dots that will appear to move across the display screen. Figure 3 illustrates this initial screen. It will be noted that somewhere in the figure there is an indicator (here shown by a line labeled "Sample Target Point") designating some specific point in the path of the dots. Shortly after the appearance of the initial screen, the dots will begin to light up in a regular pattern that will appear to the subject to be a moving light through the curvilinear pattern. This light will move rapidly, although at different speeds on each trial. The subject's task will be to monitor the progress of the light toward the indicator, and to stop it precisely at the indicator by pressing one of the buttons on the control stick. Once a trial has been completed, the subject will be given a brief period of time to inspect the results, and a new trial will begin.

4.3.2.2 Construct Analysis of this Task.

The basic skill being probed here is obviously a visually-directed motor response to a dynamic target. Obviously, it differs somewhat from standard reaction time tests that include only a discrete stimulus presentation. As such, the present task probably probes a more dynamic anticipatory skill that taps over-learned schemas for visuo-spatial timing sequences. Thus, the seemingly simple test described above may in fact probe very sophisticated and deeply learned skills.

An additional requirement that may be imposed on the subject involves dynamically changing the speed of the light as it progresses along its path. In other words, the light may be accelerating or decelerating at a constant rate. Obviously, this complicates the subject's response requirements. It adds a dimension of speed projection to the simple spatial task. The individual must not only project the moving object and its position based on constant speed, but in

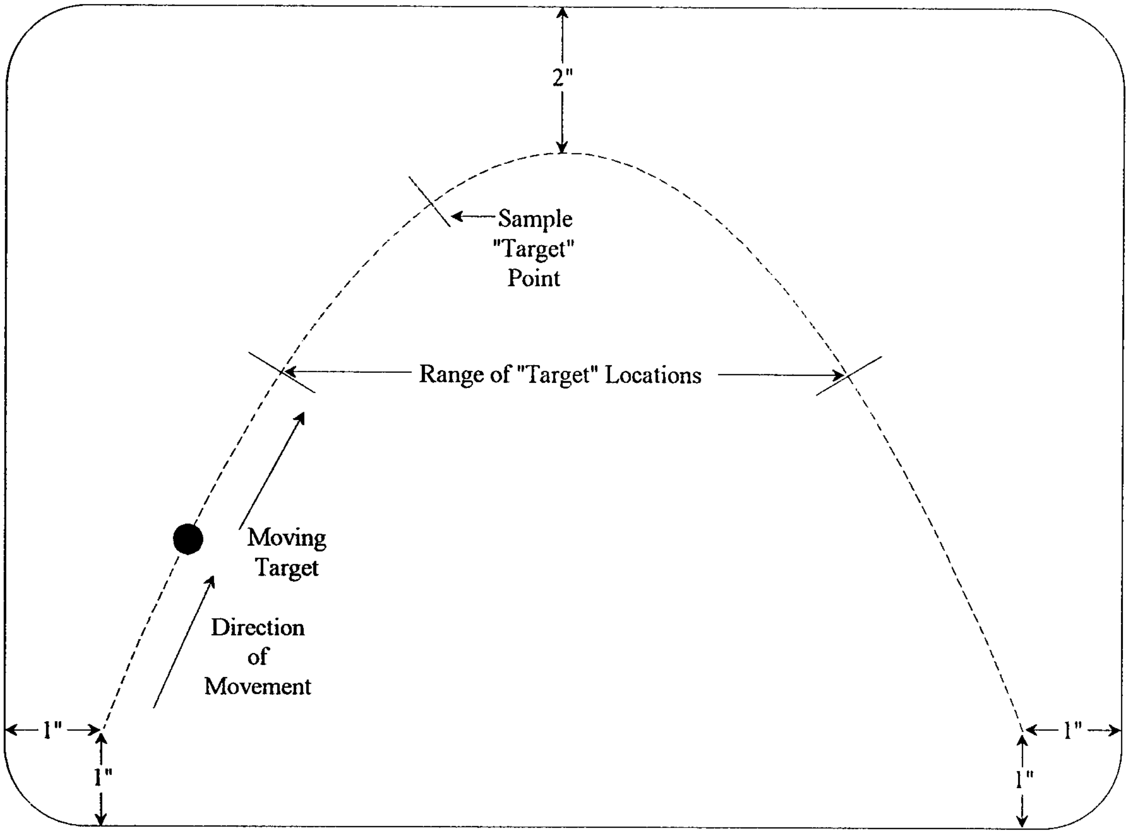

FIGURE 3. THE PRECISION TIMING TASK.

this case must perceive, process, and project the rate of change in speed. Obviously, this dimension is one that is frequently present in the operational flying environment, where an enemy may be accelerating or decelerating at the point requiring the motor response.

In summary, it would appear that this task taps an individual's reaction time capability in a dynamic visual environment. More specifically, it probes the person's ability to project the progress of a rapidly moving target, and to initiate a motor response at a precisely defined point. In its most complex form, the task also will require the individual to perceive changes in speed, and to incorporate these changes into the calculations and responses described above.

4.3.2.3 Detailed Specifications for the Task.

It is recommended that the initial screen presented to the subject contain at least 50 dots arranged in a semicircle, with its end points located one inch from the left and right side of the screen, and one inch above the bottom of the screen. The top of the circle should reach to approximately two inches below the top of the screen. The dots should be equally spaced, and no more than 1/32 inch in size.

The subject will see what appears to be a moving light that will go from left to right around the path described by the semicircle. This apparent motion will be achieved by having each dot position enlarge to twice its size in a rapidly moving sequence. In other words, starting from the extreme left hand side of the semicircle, each dot will appear to "light up" or brighten in sequence until the sequence ends at the extreme right hand dot of the semicircle.

The speed at that the light moves around the semicircle will constitute one of the major independent variables in the test. The minimum speed, or the maximum inter-flash interval, will be calculated based on the intensity of the screen ultimately used. This minimum speed will be the slowest interval that will produce the phi phenomenon. It would be expected that this would be in the range of 750 milliseconds. This would yield a trial with a maximum length of approximately 2.5 seconds. The fastest speed at that the apparent light will move will also be determined during Phase II. However, it will be defined here as the fastest achievable speed at that subjects can still give reliable responses. It would be estimated that this fastest achievable speed will be in the range of 40 to 80 milliseconds.

To achieve the capability of having the speed of the moving light change, the basic program directing the interflash intervals will permit specification of an acceleration or deceleration factor. The default option will be the constant speed. Again, it will be necessary to determine experimentally, during Phase II, what types of acceleration changes subjects can respond to reliably. However, it would not be expected that, given the difficulty of the task, a wide range of such acceleration variation will be able to be used. For purposes of this test, it will be sufficient to have a few types of acceleration variations that are easily recognized by the subject under nominal conditions.

The indicator that will tell the subject where the light should be stopped may be located anywhere within the center 1/2 of the semicircle (indicated in figure 3 by the "Range of Target Locations"). Eventually, it would be expected that a standardized protocol will be developed specifying exactly what positions should be utilized. In its most generic form, however, this parameter will be left to the experimenter's discretion. In other words, the experimenter will be able to designate exactly where he or she would like the indicator to be placed. The indicator itself will consist of a 1/2 inch line that will fall between two adjacent dots.

The testing sequence will consist of a minimum 30 trials, based on the need for reliability in this type of measure. As always, more trials are desirable, and there will be no limit on the number of trials an individual experimenter might utilize. An individual trial might last between .75 and 2.5 seconds. The intertrial interval should be constant, and should be at least 1.5 seconds long. Thus, a minimal test (30 trials) would be expected to last approximately 1.5 minutes.

Two basic types of scores will be obtained in this test. The simplest score will consist of the absolute distance error between the indicator and where the subject actually stopped the lights. This will be an actual distance measure in that distances between the dots will also be calculated. Obviously, it will be necessary to indicate whether the error was short or long (positive or negative). The summary statistics for this measure will be the mean arithmetic error, the mean absolute error, the standard deviation of both of these, and the RMS error. These will be calculated over all trials.

The second type of measure is based on the fact that there will be different levels of difficulty in many of the trials. Slower moving lights should reasonably be easier to project accurately. Although this feature of the test adds some degree of flexibility to the experimenter's analysis options (e.g., it may be found that certain stresses affect only high speed light perception and timing), it makes interpretation of the simple statistics described above somewhat problematical.

To alleviate the situation, a statistic based on time will also be calculated. In this, the subject's distance error on any given trial will be converted to the amount of time by that he or she was in error. In other words, if the subject was two centimeters short of the target, and if the light was "moving" at the rate of ten centimeters per second, the subject should have waited .2 seconds longer. This case would yield a score of .2 seconds. In another case, if the subject was again two centimeters from the target, but the light was "moving" at the rate of five centimeters per second (i.e., slower than the first case) the subject's error time was .4 seconds. This scoring system therefore accounts for the fact that the second task was presumably easier than the first. Therefore the same distance error should result in a poorer score in the second case than in the first. This calculation will be carried out automatically, and the same statistics calculated above will be calculated for this type of data.

4.3.2.4 Potential Difficulties with this Task.

The major problem foreseeable with this task is the one noted above. The difficulty level of each trial will vary as a function of the position of the indicator, and the speed of the apparent motion. Again, these independent variables will have to be controlled in any rigid experimentation. It should be reasonably easy to quantify the difficulty level of various trials by simply inspecting the total amount of time the subject has to make a decision, relative to the speed of the moving light. However, it will still remain for the experimenter to assure that these factors are well controlled.

There is some concern that the use of a simple semi-circular curve will not provide enough time or distance the test to give a good range of performance measures. Stated differently, the curve shown in figure 3 may be too short, and subjects might "learn" responses to different positions of the target line. Again, this will only be able to be determined experimentally during Phase II. However, if it proves to be a problem, one alternative is to replace the semi-circular pattern in figure 3 with a "corkscrew" pattern that would yield a much larger length of line for the moving light to traverse.

One other factor that should be of some concern involves whether or not there are differing difficulty levels associated with a light that is moving in an upward direction or one that is moving in a downward direction. In other words, is upward moving perception better or worse than downward moving perception? While there are no data that could help answer this question, it certainly should be considered either in the experimental design of the test scenarios, or experimentally in its own right.

4.3.3 Test 3. Motion Inference.

4.3.3.1 General description of the Procedure.

In the combat situation, there are instances where the individual must perceive and register the motion of an enemy or another object, then must turn away from direct visual perception of the object briefly, but still process the absent object's motion in order to know where and when it should be when it reappears. This can be a relatively brief interval, such as when an enemy aircraft passes below you, or a reasonably longer period, such as when the pilot must attend to a task in the cockpit while attempting to rendezvous with another aircraft. In such cases, the individual must infer motion based on a previous perception of motion. This must frequently be done while other tasks are being performed.

Although this is a situation that occurs with reasonable frequency in flight, it is not an easy one to quantify. Attempts to do so have essentially utilized a moving target that, at some designated point, either disappeared or was hidden from view by an obstacle. The subject had to estimate when the object, moving at a constant rate of speed, would reappear from behind the obstacle, or would "hit a target" (Weltin, Broach, Goldbach, and O'Donnell, 1992). This approach probably constitutes a good measure of the basic skill. However, it allows the subject to focus completely on this one task, and subjects sometime develop "strategies" that confound the measure (e.g., they sometimes "count beats" or sing music in time with the motion). This confounds interpretation because it is not usually the way the task has to be done in the real world. In those situations, the individual is frequently preoccupied with at least one other task while making these inferences. Therefore, it would be desirable to incorporate some form of distracting task during the period of time that the subject is making inferences.

To address this type of performance requirement, the A-PASS test will utilize a task in that the subject will see a moving light traversing a curved path (as in Task 2 above). This task is shown schematically in figure 4. The light would "go out" approximately half way through the curve. The subject's task would be to determine when the light, moving at a constant speed, would have reached the end of the curved path. In other words, the subject must infer how long the light would take to reach the end of its path. The response required will be a button press at the point at where the subject believes the light would have reached the path.

The distracting task will be a simple "semantic" one. When the light goes out, a series of four letters of the alphabet will appear on the screen. The subject must immediately decide whether any of the letters are vowels. In other words, this interpolated task will act as a distracter to the subject in estimating the inferred motion. In this way, the subject will be precluded from using methods such as counting, tapping, or singing to infer the motion. It also simulates more closely the tasks required in the real world.

4.3.3.2 Construct Analysis of this Task.

The primary skill that appears to be involved in this task is an intrinsic timing ability, similar to what has been called "time estimation". Time estimation is a fairly well studied skill in the performance literature, and has been shown to decay under several kinds of stressors that might affect the nervous system, particularly higher cortical functions (e.g., carbon monoxide, sleep deprivation, anoxia). However, its physiological basis is not well established, and to our knowledge time estimation has never been used experimentally in centrifuge research. Still, it is reasonable to expect that this task will be sensitive to the kinds of stress experienced in acceleration environments, where cortical blood flow can be compromised.

The present task differs from typical time estimation tasks, in that the subject simply counts seconds internally. In the present task, the speed of the moving target establishes a time/distance "norm" for the individual. Thus, this represents much more of a "novel" situation for the person. It forces the subject to consider an additional dimension (speed), instead of simply using the time reference. It is likely, therefore, that this task requires even higher cortical activity than simple time estimation tasks, and should be more sensitive to cortical insult than those tasks.

It should be noted that the first task of the subject is to establish a "rate" of movement for the target. This rate is a ratio of distance traversed per unit time. This is a skill that is constantly used in many situations. However, this task requires that the subject establish the rate extremely quickly, that makes it much more like the task of the fighter pilot. Once established through observation, this

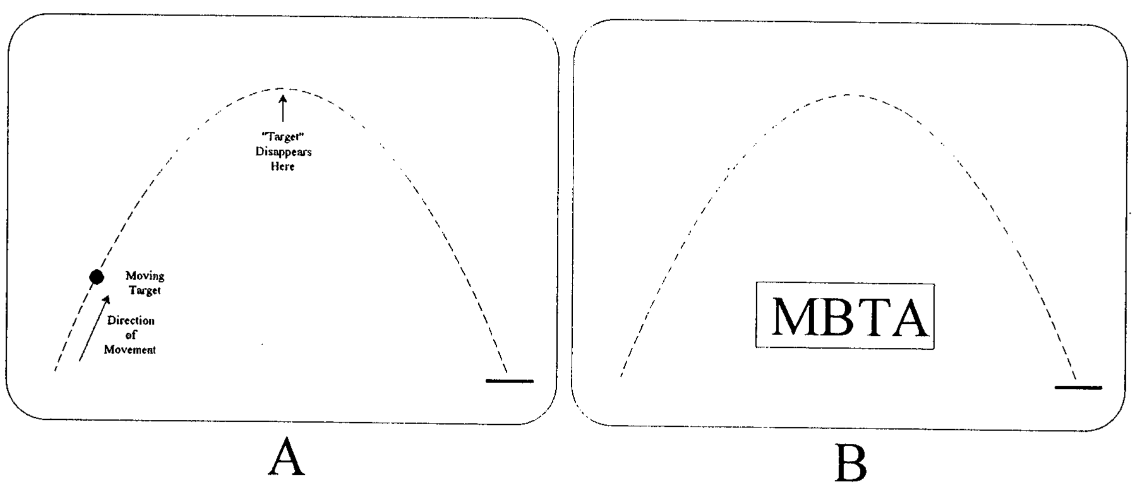

FIGURE 4. THE MOTION INFERENCE TEST.

rate must be maintained internally (and in the presence of a distracter) for some period of time. Again, while not well defined, this is a skill that is used in many real-world activities, from driving an automobile to sports activity, and is easily recognized in the activities of fighter pilots..

4.3.3.3 Detailed Description of the Task.

This task will utilize the same basic stimulus environment as the precision timing task described above. In other words, the path traversed by the "light" will be the same as in that task. The subject will initially see the dots describing the path that the light will take. The "light" will consist of a brightening and enlargement of sequential dots at a rate sufficient to generate the phi phenomenon. The light will appear to move at a constant speed for each trial. However, the speed will differ from trial to trial. The range of speeds at that the light will appear to travel will, of course, be optimized experimentally during Phase II. Nominally, speeds ranging from .5 to 2.5 inches per second will be used in initial trials. Again, Figure 4A illustrates the appearance of the screen to the subject, including the moving light.

The apparent light will go out as soon as it reaches the center of the semicircle. Immediately, four alphabet characters will appear at the center of the arc described by the semicircle, as shown in figure 4B. These will subtend approximately .5 minutes of visual angle, in order to be large enough not to be affected by small visual acuity changes. They will be located not more than .5 inches from each other in a straight line. The subject's task will be to indicate, as quickly as possible, whether the series contains a vowel or not. This will be done with the left hand by pushing the throttle either forward or backward to indicate the proper response. Once this is done, the subject will then estimate when the light would have reached the extreme right-hand end of the semicircle, and will indicate this by pressing a "fire" button on the stick.

The selection and distribution of letters for this subtask is, of course, critical. Different letters will be used on each trial. However, certain parameters will be established to ensure that, within each series of 32 trials, the following conditions will occur:

a. There will be a 50-50 split in the probability of a vowel appearing in the sequence.
b. Only one vowel per series will be present.
c. The vowel will appear in each position an equal number of times.
d. The same vowel will not appear in two contiguous trials more than twice.
e. No "run" of more than four trials consecutively will be permitted for either vowel or non-vowel conditions. Beyond this requirement, no randomization formula will be used.

To achieve the above control over the parameters of the presentation of the task, a "canned" set of 32-trial series will be produced. At least 20 such series will be developed and available for presentation to the subject.

It is estimated that a single trial for this task will last between two and four seconds. NTI recommends at least a three second intertrial interval to permit subject feedback and rest. Thus, each trial will occupy between five and seven seconds. Nominally, at least 32 trials should be given in a test session in order to minimize variability. Therefore, it would be estimated that the test overall might last, on the average, between 3 and 5 minutes.

The type of data produced by this task are very similar to that produced by the precision timing task. Error scores will be available in the form of either distance or time between the actual target point and at the point that the subject responded. Therefore, the data will be treated in the same way as described for the precision timing task.

4.3.3.4 Potential Difficulties with this Task.

A unique confounding factor in this test is the interpolated task. This is deliberately intended to be an extremely easy task, and therefore should not show much variation. However, in some situations, it might be necessary to consider performance on this "secondary" task as a potential confound to the interpretation of primary task performance. Obviously, if the subject ignores the secondary task, primary task performance will be improved. Conversely, if the subject finds this task particularly difficult after an unusual stressor (one that affects "semantic" memory) primary task performance could be degraded. Finally, inconsistency on the part of the subject with respect to giving attention to the secondary task could introduce great variability into the data.

Obviously, these difficulties are the same as those faced in any multi-tasking situation. Only if there is some anomaly in the secondary task will we need to consider modulating our interpretation of the primary task performance. In view of the this, NTI recommends that the semantic secondary task be scored in terms of both response time and percent correct. During Phase II we will establish experimentally the range of "normal" performance on this secondary task. This then will be incorporated into the scoring system for this task.

If an individual falls within this "normal range", nothing will be done further to modulate the performance score of the person on the primary task. However, if the person falls outside this normal range, his or her performance score on the primary task will be changed, depending on the pattern of performance between the two tasks. Again, this will only be determined definitively after experimental data are available. However, certain tentative speculations can be advanced regarding the meaning of such interactions. For instance, if the secondary task performance falls below its norm, or below the subject's usual performance levels, then primary task performance would be inspected. If it significantly improved, the assumption might be that the person had shifted attention to the primary task at the expense of the secondary. In that case, either the trial would be ignored, or a penalty would be assessed against primary task performance (i.e., it would be considered to be more degraded than it appeared to be). On the other hand, if the primary task performance also degraded, or even stayed the same, it might be assumed that the decrement in secondary task performance represents a true performance effect, and this might then be interpretable in its own right.

4.3.4 Test 4. ILS Landing Simulation.

4.3.4.1 General Description of the Procedure.

There was virtually unanimous agreement among pilot consultants and Government pilots meeting during Phase I that one of the most challenging tasks for a pilot, and one that provides a benchmark for many less challenging tasks, is an ILS landing. In this, the pilot must monitor and integrate a number of data inputs while simultaneously performing a reasonably difficult psychomotor tracking task based on visual input. The actual operational task also involves elements of decision making, in that the pilot must decide whether certain critical conditions exist that might justify an aborted landing or other activities. The ILS approach therefore appears extremely attractive as one of the critical tasks in the A-PASS system.

In this instance, NTI decided that it was advisable to create a part-task simulation that was reasonably close to the real world environment. In other words, actual HUD symbology, as well as reasonably high fidelity flight dynamics were desirable. This approach was taken because the ILS approach task is, in itself, a reasonably pure measure of certain critical performance skills. The recommended approach includes a moderately high fidelity ILS approach screen modeled after the F-16 aircraft. In addition, strip indicators for both altitude and airspeed will be used, and all of these will respond accurately to control inputs, based on an elaborate aerodynamic model of the F-16 aircraft. The subject will be required to carry out a standard approach down to a point slightly below "decision height". The screen viewed by the subject is shown schematically in figure 5 on the next page.

The task for the subject will be to actually "fly" an aircraft simulation from a designated point to an appropriate decision height point. In reality, this means that the problem will be presented to the subject essentially as a tracking problem. A standard F-16 display will indicate to the subject whether he or she is on the "glide path". This glide path indicator is illustrated in figure 5 as the small circle and vertical/horizontal crosshairs. The crosshairs on this indicator tell the pilot whether he or she is above or below the required glide path.

In addition to the glide path indicator, two additional indicators will be given to the subject. One will indicate the altitude of the aircraft, and the other will indicate the airspeed. These will be tied to the individual's joystick and throttle, using dynamics of the F-16 aircraft. As the individual flies the aircraft toward the decision point, these indicators will change appropriately. As the individual descends, the glide slope indicator will move up/down or right/left. This will indicate to the subject that an appropriate correction must be made with the stick. The airspeed and altitude will also move appropriately with the individual's stick and throttle inputs. The goal of the subject is to maintain the

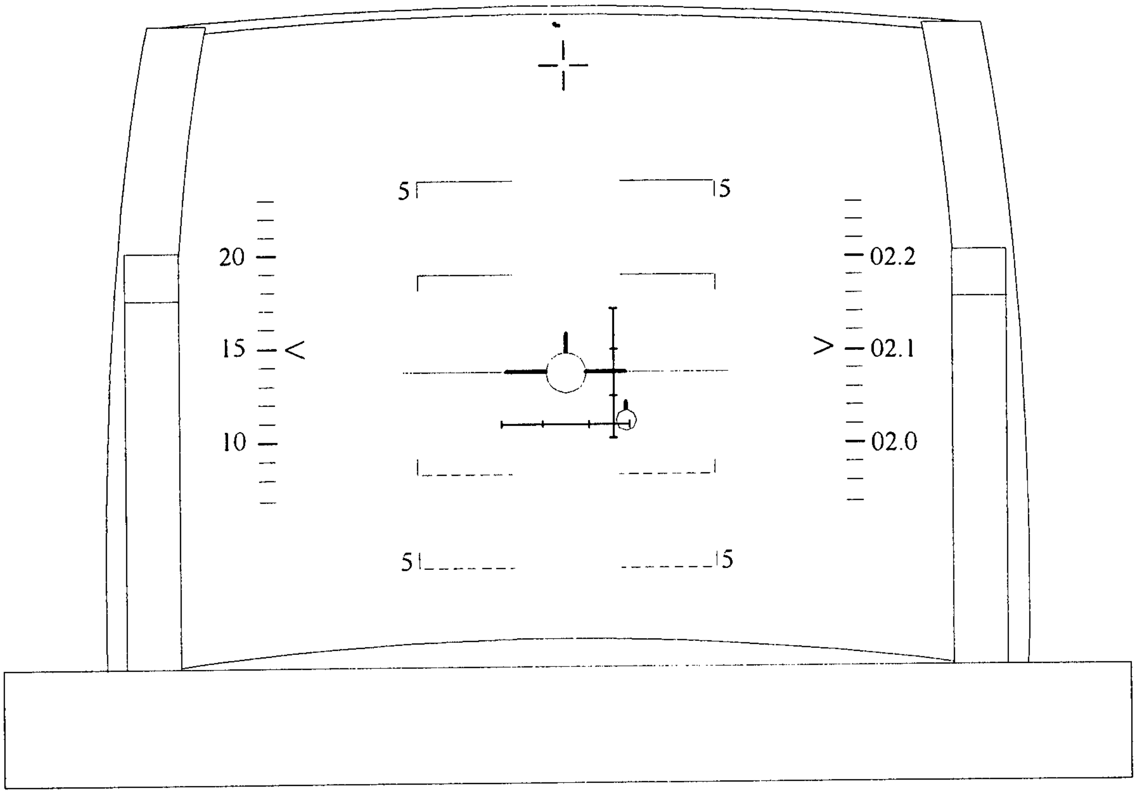

FIGURE 5. THE ILS APPROACH TASK.

circle at the center of the glide slope indicator. This will represent an appropriate rate of descent and position in three dimensions.

The airspeed indicator must be monitored to assure that a correct directed airspeed is maintained for various points in the approach. The altitude indicator must be inspected periodically because a critical "decision height" will have been instructed. If the situation at any point in this approach falls within certain unacceptable parameters, the individual will have been instructed to abort the landing or continue the ILS approach. Thus, the subject will essentially perceive a tracking task with additional monitoring requirements.

4.3.4.2 Construct Analysis of this Task.

This procedure essentially is a very complex tracking task with multi-tasking elements. The tracking portion of the task is not extremely difficult in itself, except that the "criticality" of the control movements changes as the task proceeds. This is similar to the changing demands placed by a critically unstable tracking task (i.e., the task continues to get more difficult over time), but without the instability. In essence, if tracking was all the individual had to do (and provided excessive wind gusts were not employed), the trained subject should have little difficulty with the task.

The difficulty is introduced by the fact that both airspeed and altitude must be monitored simultaneously with tracking. Although these can be considered probes of the person's monitoring skill, neither of these tasks would be extremely difficult in themselves to the trained subject. However, the combination of all three tasks introduces the need for the subject to divide attention (attention allocation), to integrate all three types of information, and to make control decisions based on that integration. The overall task, then, probably taps attention and multiplexing skills in the subject more than any of the specific skills tapped by the individual tasks. This makes it most appropriate for inferring the pilot's ability in a variety of multi-tasking requirements of flight, and accounts for the fact that the pilot consultants considering this an excellent example of a "critical" task.

If wind gusts are considered, the tracking task itself becomes more difficult. By manipulating the amount of gusting, the experimenter can drive the subject's attention away from the two monitoring tasks and make the overall multiplexing task more difficult. In this sense, this task provides a considerable range of options to the experimenter. Attention allocation and multiplexing skills can be tested at low stress levels, or at the outer limits of the person's ability to control the aircraft.

4.3.4.3 Detailed Description of the Task.

Programming of this task will be reasonably complex. It is desirable to have control/response characteristics in the altitude, airspeed, and ILS indicator. To this end, NTI proposes that the F-16 aeromodel developed by Dr. Brian Stevens (Georgia Tech Research Institute) be incorporated into this task. In other words, that portion of the aeromodel that addresses F-16 reactions to control inputs at the appropriate altitudes and airspeed will constitute the basic computer structure of this task.

In present case, the ILS indicator will subtend approximately four degrees of visual angle. The moving circle within the ILS indicator, as well as the crosshairs and strip indicators, will be consistent with the size of an actual F-16 display. The altitude indicator will be located to the right of the ILS indicator, and the airspeed indicator will be located to the left, as shown in figure 5. Again, the goal will be to present the individual information bars as close to the way they are presented in an F-16 aircraft as possible. Similarly, movement of the strip should resemble actual F-16 indicator movement.

A trial in this test will consist of requiring the subject to "fly" the aircraft from a point about 7 miles away from the touchdown point to a decision height altitude near the touchdown. The subject's task will be to intercept and remain on the glide slope by keeping the target centered in the crosshairs. This will require, of course, coordinating speed and altitude loss. A trial should last three minutes nominally.

The pre-set 3-degree glideslope will simulate a standard straight-in approach requiring the subject to descend at approximately 700 - 800 ft. per minute to maintain the glidepath. The HUD ILS display will not include the F-16 command steering symbol in order to simulate the more demanding "raw data" or "needles only" approach. Simulated angular deviations of the aircraft from the ILS course and glidepath will be indicated by the vertical and course deviation bars. Simulated aircraft control responses will be representative of the F-16.

The relative difficulty of each trial will be determined by the amount of wind and gust programmed for that trial. At least five difficulty levels will be available to the researcher. In the nominal case, there will be no disturbance on the aircraft. In the extreme difficulty case, (designed to maximize the chance that the aircraft will be out of allowable standards at decision height) wind will be high and gusts will be introduced (wind shear) close to decision height. The remaining three difficulty levels will all be designed to make the task able to be done, but to introduce a wide range of difficulty in achieving a good score.

It should also be remembered that the nature of the glide path indicator makes the task more difficult the closer one comes to decision height. In effect, the aircraft can tolerate a much wider range of variation, either lateral or vertical, when it is far from the landing point. As it gets closer, this allowable range decreases, and small movements of the aircraft have a higher effect on indicated error.

4.3.4.4 Potential Difficulties with this Task.

This phenomenon noted above introduces some difficulties into the scoring system for this test. Essentially, a deviation that is meaningless 7 miles from the touchdown point might be catastrophic at a quarter of a mile. Therefore, tracking error of actual deviations alone will not provide a meaningful metric for this test. In order to have the metric most efficiently serve as an input into the computer model, the "criticality" of an error at various points in the glide slope should be incorporated into the measure.

Since the ILS course deviation "needles" display only angular deviation from the desired course and glidepath, the actual lateral and altitude error is directly dependent on the distance from the touchdown point. A "1-dot" deviation at 7 nm from the landing is roughly seven times the lateral or vertical error of the same "1-dot" deviation at one mile from touchdown. Obviously, as the aircraft nears the landing point, the needles appear to become more sensitive, and the task of maintaining the course and glidepath becomes more difficult. Since given actual deviations become more critical (i.e., have more impact on the safety and effectiveness of the approach) the closer the aircraft is to the touchdown point, the scores might be weighted inversely by the range to touchdown. Effectively, therefore, the subject may be scored on operational effectiveness and safety by using the weighted error of the deviation bars.

Since the difficulty of maintaining a prescribed airspeed does not change during the approach, RMS airspeed deviations from the target value may be scored directly. Like position deviations, however, airspeed becomes more critical as the landing point is approached, so some scaling based on the inverse of range to touchdown may also be appropriate here.

4.3.5 Test 5. Pitch/Roll Capture Task.

4.3.5.1 General Description of the Procedure.

One of the critical skills described by our pilot consultants involved rapidly positioning the aircraft in order to move the bandit into a specific location relative to your aircraft's aiming devices. Essentially, the pilot must recognize the bandit's relative position, and then must make a rapid correction in his or her own position and attitude in order to gain the proper aiming advantage. This rapid maneuver might be in the vertical relative to the aircraft ("pitch capture"), or laterally (roll capture). These terms refer to the required control input, pitch or roll, that will bring the bandit into the desired position. Delays or errors in doing so will obviously have an impact on the outcome of the air engagement. On the other hand, if the pilot demonstrates a normal ability to carry out this type of task, it should be possible to assume that the person will be able to carry out other, less rapid "capture" tasks (e.g., formation flight).

To probe the skills required for this type of task, the A-PASS battery will contain a medium-fidelity simulation of the actual pitch and/or capture as it would be carried out in an F-16 aircraft (see figure 6). The subject will see a crude front cockpit simulation, and will be required to be performing a routine, easy cockpit task (e.g., flying straight and level, or adjusting a radio frequency). At some random time during the trial, a target will appear in some location around the cockpit field of view. The subject's task will be simply to move the target as rapidly as possible into an instructed firing position, using the control stick.

4.3.5.2 Construct Analysis of this Test.

The basic skills required by this task consist of a rapid determination of the movement required, followed by very rapid and precise visually-controlled motor response. In essence, the subject must maintain a level of alertness for an expected situation while performing a simple primary task. When the situation actually occurs, a demanding, precision response must be initiated. The skills required here are somewhat different from the dynamic reaction time measured by the precision timing test (No. 2) described above, or by any simple reaction time test. In the present case, the response involves a rapid decision as to the location of the target, a determination of the appropriate response, translation of that determination into a relatively complex control movement of the stick, and a capture of the target within rigid limits. Although the response sequence will have been over learned, at least in its gross aspect, it is exactly the type of response that could be degraded by physiological insult, especially those that might affect the blood supply to the brain.

In view of this assessment of the skills required by the this task, it would be expected that the data produced from this task will relate to any situation where the pilot must respond rapidly to the appearance of a novel situation (e.g., missile

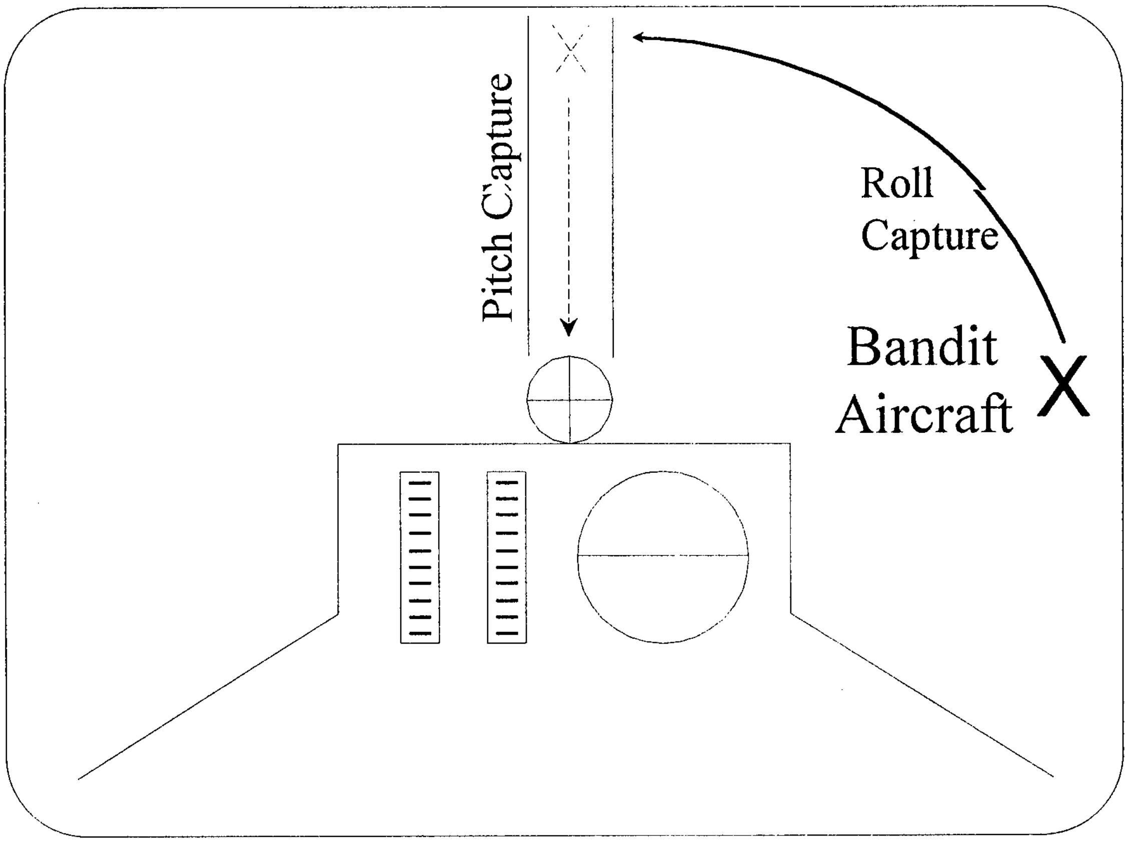

FIGURE 6. THE PITCH/ROLL CAPTURE TEST.

avoidance) or any task requiring any rapid and precise maneuvering of the aircraft.

4.3.5.3 Detailed Specifications for the Test.

The initial screen viewed by the subject will consist of a forward cockpit view as shown in figure 6. This will present sufficient information for the subject to "fly" a straight and level path. Subjects will do this for a random amount of time, up to one minute in duration. At some point during this time, a "target" will appear on the HUD. The target will appear either directly above the aiming reticule, or offset to the side. In either case, the subject's task is to make a control movement of the stick, as quickly as possible, that brings the target into the aiming reticule. In the case where the target is directly above the reticule, this would involve pulling the stick back (pitch capture) to simulate raising the nose of the aircraft. If the target appeared directly to the side, the required response would be to "roll" the aircraft to one side or the other in order to bring it into the reticule. Any oblique appearance of the target would require first a roll maneuver, and then a pitch maneuver. In all cases, the subject will be required to press a "fire" button when he or she has achieved "capture" of the target.

At least 32 "standard" initial positions of the target will be pre-programmed for this test. These will be segregated into several levels of difficulty (depending on whether they require small of large corrections, and whether they require single or dual capture). The researcher will therefore be able to customize the level of difficulty of the tasks to suit individual research requirements, and to assure that the difficulty level of tests done at various times are equated.

A single trial in this test would be expected to last no more than one minute, with an average of slightly more than 30 seconds. Since this is a reasonably simple skill that is being tested, it would be expected that stability would be reached with one to two hours of training. Reliable results could then be expected with as few as 15 trials. Therefore, the duration of the test would be expected to be between 10 and 20 minutes. This should permit calculation of performance on this skill under a wide range of difficulty levels.

Scoring of this task is straightforward. The basic dependent variable is the time taken to "fire" - in other words, the time taken to perform the capture. Of course, it will be necessary to modulate this measure by some estimate of the accuracy of the subject when the "fire" button was pressed. This will be done in a fairly crude way in the sense that, if the target was within the reticule when the fire response was made, the time will be accepted. If the target was not within the reticule when fired upon, the task will be considered "failed" because the target was missed, and the subject will be considered "killed". Again, the time to achieve the capture will be interpreted in terms of the difficulty of initial conditions. This will provide a rich source of data input to the computer models described below.

4.3.5.4 Potential Difficulties with this Task.

This is one of the more direct measurements obtained in the A-PASS battery, and therefore has fewer interpretational difficulties associated with it. It is a reasonably direct analogue of a critical task in the cockpit and as such, provides a direct input into many aspects of missions. The major potential problem again revolves around assuring that the difficulty level of trials is appropriately recognized, and that data from one level of difficulty are not used to infer performance at another level. For instance, if no decrement is found in performance for a simple pitch capture task, one should not infer that there will be no difficulty in a complex pitch/roll capture task. Appropriate experimental design should totally negate this problem.

4.3.6 Test 6. The Gunsight Tracking Task.

4.3.6.1 General Description of the Procedure.

A relatively high fidelity simulation of the gunsight tracking task carried out by the pilot was recommended by the consultant pilots. A simplified version of the display seen by an F-16 pilot during a ground attack will be presented to the subject (figure 7). The gunsight will be controlled by the subject, using the stick

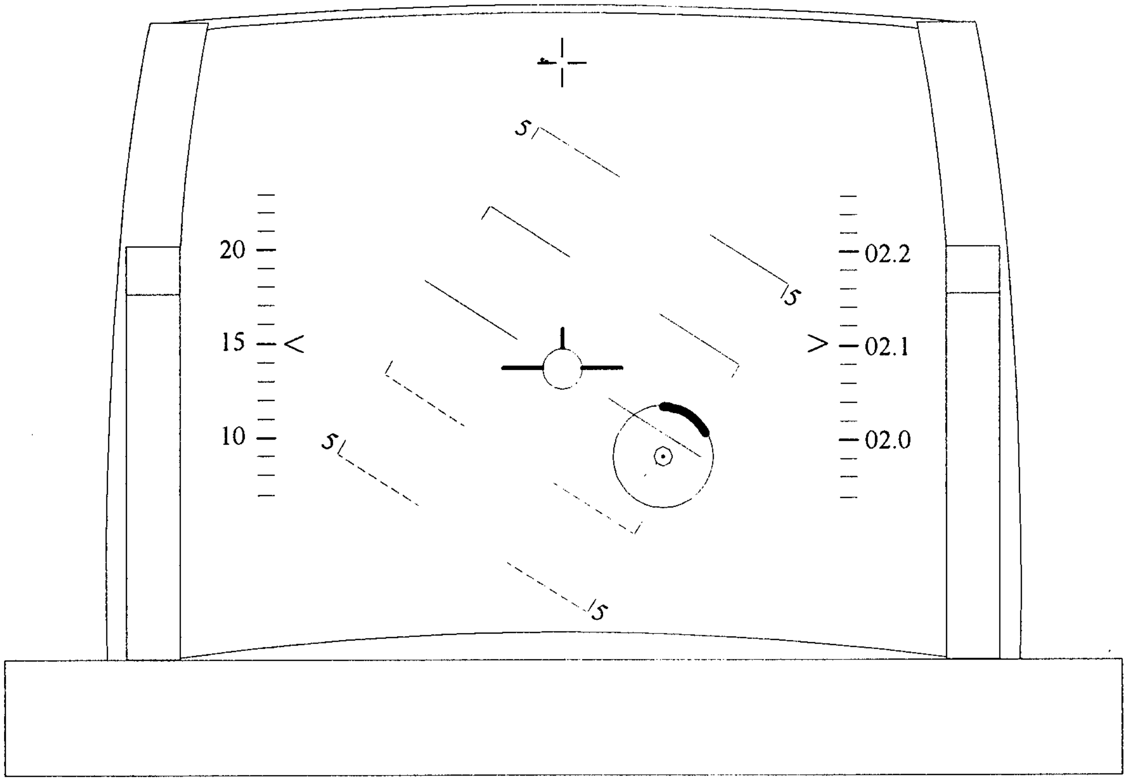

FIGURE 7. THE GUNSIGHT TRACKING TASK.

of the simulator, thus simulating moving one's own aircraft to bring it into firing range. The control functions that drive the apparent movement of the target will be precisely determined, yielding a large number of precision metrics dealing with the pilot's ability to capture the target. Obviously, measures of the individual's speed and accuracy on this task could be used with or without a mission model. In other words, this is a task in that the data would provide almost direct OMI metrics to the field commander.

4.3.6.2 Detailed Description of the Test.

The stimulus presentation to the subject for this test will be very similar to that used in the ILS landing test described above, except that the symbology reflecting the gunsight pipper will replace the ILS glide path indicator. This presentation is shown in Figure 7. The subject will control the pipper in the same way it would be done in an F-16 aircraft. In other words, the subject's task will be to maneuver the aircraft in such a way that the weapons delivery would be targeted on a specific point. This maneuvering will be carried out through use of the stick and rudder combination. The subject's task will be to bring a moving "target" into the pipper, and then to "fire" the weapon precisely at the point where the target and pipper meet.

Specifically, the target to be used for this test will be identical to the target symbol used in the F-16 aircraft. The target may appear in any position on the HUD (i.e., above, below, or to either side of the pipper). Target motion will be determined by the airspeed, angle of attack, and heading of the subject's aircraft. In other words, if the aircraft angle is too low, the target will appear to move faster. If the aircraft angle is too steep, the target may remain stationary, or move very slowly. Again, these values will be determined based on the dynamics and the configuration of the F-16 and its weapons system. The subject's task, of course, is to join the target and the pipper as quickly as possible by maneuvering the aircraft in the appropriate direction. In this sense, the tracking task represented by this test is one in that the target's movement is totally determined by the subject. This, of course, is precisely what occurs in a gunsight tracking condition, where the target is stationary and the aircraft is moving.

The duration of each tracking trial will be determined by the initial conditions and the subject's responses. Realistically, it would not be expected that a trial would last more than five seconds, since this is nominal for a realistic gunsight tracking task. With adequate training and appropriate control of the stimulus materials, it would be anticipated that a stable measure could be obtained with as few as 12 stimulus presentations. Therefore, with rests, the duration of the test could be as little as two minutes.

Scoring of this task will be straightforward. The time taken to link-up the target and the pipper, and the accuracy of the release response by the subject will be scored. Time will be measured from the start of the trial to the point of nearest approach of the target to the pipper. Accuracy will be measured by the absolute error distance between the target and pipper at the time of weapons release. A third measure, the time from stimulus presentation to weapons release, will also be calculated. These metrics will provide data on the efficiency of the tracking, as well as the accuracy of the response or weapons delivery. It should be noted that, in these forms, the data provide directly useable input to the operational commander. Any decrement in weapons release point from this test could be directly translated into CEP or damage assessment.

4.3.6.3 Potential Problems with this Procedure.

Other than the fact that this will be a rather difficult test to program, there are few unique problems associated with its administration. It is a direct analog of a real world task. There are, however, the same concerns that have been described for early tests. First, not all trials will be equally difficult, since the initial position of the target will determine how much maneuvering the subject will have to carry out in order to effect a successful weapons release. Therefore, as in all of the tests described above, extreme care will have to be taken to equate the stimulus conditions across experimental runs in order to assure that the data will not become biased.

4.3.7 Test 7. Peripheral Monitoring Test.

4.3.7.1 General Description of the Procedure.

In the course of both routine and specialized operations, the pilot is constantly required to monitor many things at once. Most of the tests in this battery concentrate on what has been called focal attention, in that the pilot must foveate the stimulus materials. It is desirable, therefore, to add a test procedure that assesses the individual's peripheral visual capabilities.

This is particularly true in the acceleration environment, where one is interested in the loss of peripheral vision under high G_Z forces. The loss of blood from the retina and eventually from the brain due to such forces results in early degradation of the peripheral monitoring capability of the individual. This fact has long been recognized, and has resulted in the peripheral lights test becoming one of the more standard assessment techniques in acceleration research (Coburn, 1970; Fong, 1992). It constitutes a reliable and extremely valid indicator of the person's capability to perform. However, the peripheral lights test is not a performance test *per se*. It simply indicates when peripheral lights disappear as a function of loss of blood in the retina. What is needed is a peripheral perception test that indexes not only the loss of lights but a phenomenon that presumably will occur earlier, namely the loss of information processing ability from peripheral stimuli.

To address this need, the A-PASS test battery will contain a peripheral monitoring task along with a standard peripheral lights test. The apparatus for this test will be specially constructed for installation into the centrifuge by Government personnel or contractors. This apparatus is shown schematically in Figure 8. It will consist of a relatively standard semi-circular light bar with individually illuminated lights. One of the most basic tasks to be performed by the subject would be to indicate when specific peripheral lights disappear. Essentially, this portion of the test will be identical to the procedures currently used in the Centrifuge.

The novel portion of the test will involve a series of displays that will be mounted on vertical bars, and will be able to be positioned at various peripheral locations by moving them on the semi-circular light bar. These displays will be capable of indicating a variety of types of information common to the cockpit environment. Obviously, these types of information will be types that are commonly presented in the periphery of the pilot's vision.

The subject's task in this portion of the test will be to monitor the peripheral information, while carrying out some function that requires that foveal vision be occupied. In other words, this foveal task will insure that the pilot must acquire the peripheral information non-foveally. The central task will be one that is reasonably easy to perform, but that requires constant attention.

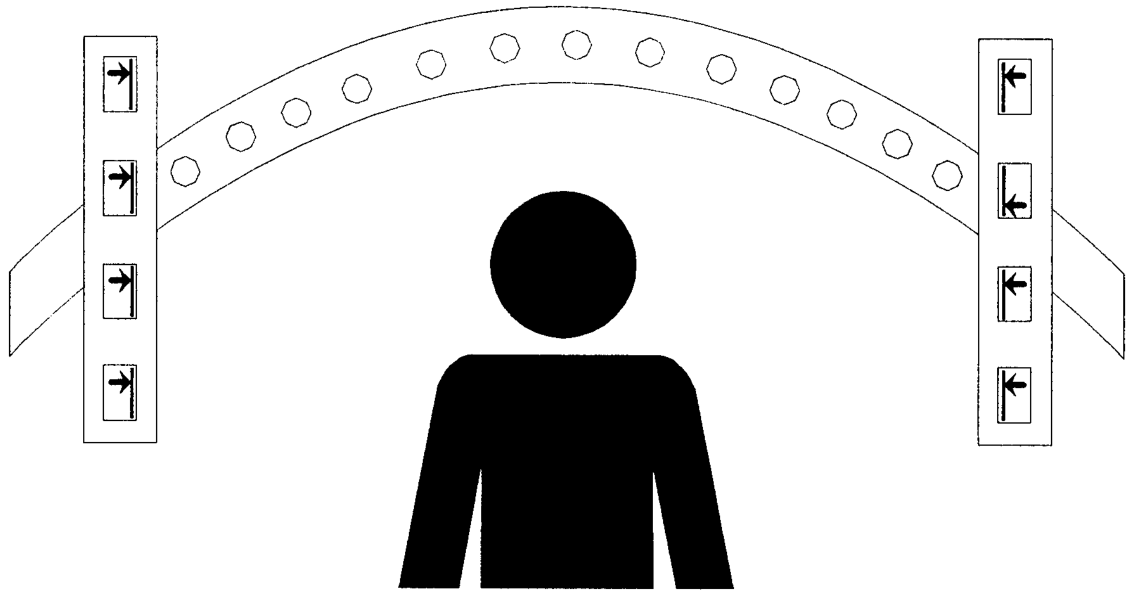

FIGURE 8. THE PERIPHERAL MONITORING TEST.

4.3.7.2 Construct Analysis of this Task.

The peripheral monitoring of information is reasonably well understood in terms of its physiology and its phenomenology. The peripheral retina consists predominately of rod receptors. While these receptors are extremely sensitive to light, their retinal interconnections yield extremely poor acuity. However, they are exceptionally well suited to monitoring movement. For this reason, peripheral displays typically utilize some form of gross movement rather than small, precise display of static information. Thus, it would be anticipated that this test would be extremely sensitive to loss of ability to monitor the peripheral movement that is characteristic of many aircraft displays.

The skills likely to be probed by this task involve principally the ability to detect motion (or the lack of it) in peripheral displays. Depending on the type of displays used, changes in this skill will be readily interpretable in terms of the kinds of errors that a pilot might make, and the information that would be compromised as a result.

4.3.7.3 Detailed Description of the Test.

The exact dimensions of the apparatus to be constructed for this task will depend, of course, on the specifications and limitations imposed by Government personnel as a result of the centrifuge requirements. However some approximations can be made at the present time. It can be estimated that the semi-circular light bar will have a radius of roughly 3 feet. It would be anticipated that the bar itself would be approximately three inches high, and would have a "carrier bar" attached to its rear surface to provide a slide for the vertical bars. These vertical bars would be approximately two feet high by six inches wide. Four rectangular holes will be spaced equally along the height of the bar, as illustrated in Figure 8. These holes will be approximately four inches by two and half inches, or whatever size would accept standard aircraft peripheral displays (or simulated models of those displays).

It is anticipated that at least two different types of dials will be used in the present task. One will consist of circular dials with pointer indicators. The other type of dial will consist of strip indicators with a needle capable of moving up and down the dial. Eight indicators of each type will be fabricated.

These indicators will be controllable by the computer program in such a way that each dial can be independently moved. In the actual testing situation, it would be anticipated that two general types of stimulus presentation would be used. In the first, all dials on both sides of the semicircle will be placed in the same position. The subject's task will be to detect when any dial moved away from that position. In the second general condition, dials would be constantly moving, but within a "normal" range of movement. The subject's task will be to detect when any dial goes outside this normal range (e.g., when any dial indicated a danger situation). Of course, several variations in these two basic approaches are possible, and will be programmable by the experimenter. The subject's response to an "out of bounds" condition will be to perform a control movement (e.g., a button press, or a stick movement) that will cancel the out of bounds condition.

The basic measurement parameters of this task will consist of the time between the occurrence of an out of bounds condition and the subject's response. It will also be possible to specify a "time-out" value that will indicate that the subject simply did not see the anomaly. This would simulate catastrophic failure to attend to the peripheral information.

In order to assure stability in this type of task, it is anticipated that a short (e.g., 30 minute) training time will be required. The task essentially is already overlearned in that individuals readily are able to monitor peripheral information of the type to be presented here. Once stability is achieved, it would be anticipated that no more than thirty well-balanced trials will be necessary to obtain a reliable performance from the subject. The duration of each trial could be as short as ten seconds. In that case, the average time for the appearance of an out of bounds condition would be five seconds, with a range from one to ten seconds. Of course, the experimenter is free to adjust this time at will, and it would not be unrealistic to introduce individual trial times as long as one or two minutes.

The "central" foveal task to be presented to the subject can be tailored to the individual experiment. For instance, it could be a routine flying task, or even an air-to-air engagement. This would be determined by the experimenter based on the type of inference that was desired from the data. Nominally, it would be anticipated that this central task would be relatively simple and essentially non-threatening to the subject. In other words, it would be desirable to obtain a pure measure of peripheral monitoring, in most cases, where the subject was reasonably free to carry out such monitoring. Usually, the sole function of the central task will be to assure that the subject is not able to attend to the peripheral tasks with focal attention. For this purpose, a simple flight task, such as maintaining straight and level flight in a reasonable wind condition might be appropriate. In fact, any synthetic task (e.g., a tracking task or a cognitive processing task) could also be used for this purpose.

On the other hand, in some situations, the experimenter might be interested in determining peripheral processing capability when the subject is engaged in a highly demanding central task. In this case, the present case could be combined with any of the other APASS tests, or with another flight simulation task that would place higher demands on the subject's focal attention. Again, the flexibility provided here yields a great deal of information that can than be used in computer simulation models.

4.3.7.4 Potential Problems with this Task.

Since peripheral monitoring is a reasonably well understood construct, there are few experimental or theoretical difficulties encountered with this test. However, as designed here, it should be remembered that this task is a "divided attention" task.

There is a classic problem with divided attention approaches in that one is usually not certain how the individual has divided his or her attention. A decrement, therefore, may be due to either a true physiological insult, or simply to the fact that the person has tried a different division of attention. This is a universal problem with all divided attention tasks. Typically, researchers attempt to circumvent the problem by instructing the subject that one task is primary and one task is secondary. However, in the flight environment this may not be an ideal solution.

It would be hoped, in the present case, that the subjects in this type of experiment would evolve a relatively consistent pattern of divided attention, and that this pattern is reflective of the normal division of attention given by a pilot. However, in developing the protocols for this test procedure, some attention will have to be given to the question of division of attention.

4.3.8 Test 8. Unusual Attitude Recovery.

4.3.8.1 General Description of the Procedure.

Situation awareness is recognized as one of the prime determinants of successful flight. Although there are many sophisticated and subtle ways to measure the construct of situation awareness, many of these are relatively time consuming and inappropriate for situations where the experimenter is constrained by time or availability of subjects. Therefore, although it was recognized that some measure of situation awareness should be obtained in the A-PASS battery, many of the procedures that are being developed to assess situation awareness were inappropriate in the present context.

One approach, however, directly related to the ongoing development of situation awareness measures, appeared feasible in the present context. This was the assessment of the pilot's ability to respond to unusual attitudes of the aircraft.

Unusual attitudes refers to the occurrence of an aircraft position that the pilot had not anticipated based on previous data. They represent surprise conditions to the pilot, and typically must be responded to immediately to avoid catastrophic outcomes. As such, these conditions represent relatively simple but intense examples requiring the pilot to demonstrate instantaneous situation awareness. For these reasons, an unusual attitudes test was included in the APASS battery.

The measurement technique adopted in this test essentially reflects the situation in that a pilot would have HUD information indicating unusual aircraft attitudes. The F-16 HUD again was chosen as the basic display for this test. The individual will be presented with the essential information from the HUD indicating that the aircraft had assumed an attitude that, 1) could not have been anticipated by the pilot, 2) represented a dangerous condition of the aircraft. The individual is required to take immediate and appropriate action, through use of the control inputs, to counter the unusual attitude.

In the present case, the display presented to the individual will be identical to that used in several of the tasks described above. It will be a moderately high fidelity simulation of the F-16 HUD (figure 9). In this case, the HUD will represent simply the information necessary for control of the aircraft (i.e., the gun side tracking information and the ILS information will not be presented). A HUD will appear suddenly at some point during the subject's trial, and will represent a situation that would have been unexpected by the pilot, and that requires immediate remedial action. The subject's task, of course, is to correct the unusual attitude and to bring the aircraft into a stable set of flight parameters. The adequacy of the subject's response in carrying out these corrective actions will constitute the basic metrics of this task.

4.3.8.2 Construct Analysis of this Task.

The variety of possible corrective actions that a subject can make in such a situation precludes in any detailed analysis of the specific skills involved in this task (see discussion of potential problems below). Therefore, this test is not viewed as one that probes individual skills of the pilot (although a detailed and somewhat tortuous analysis might reveal such skills). Instead, this task is viewed as a generic indicator of the subject's situation awareness, defined as the ability to immediately perceive the status of the environment and of his or her own ship in a particular context (McMillan, et al., 1994). As such, this task should serve as an adequate modulator of the pilot's performance in a wide variety of situations. If situation awareness is degraded, this will provide a meaningful data point (and distribution) that can serve as input to many operational tasks.

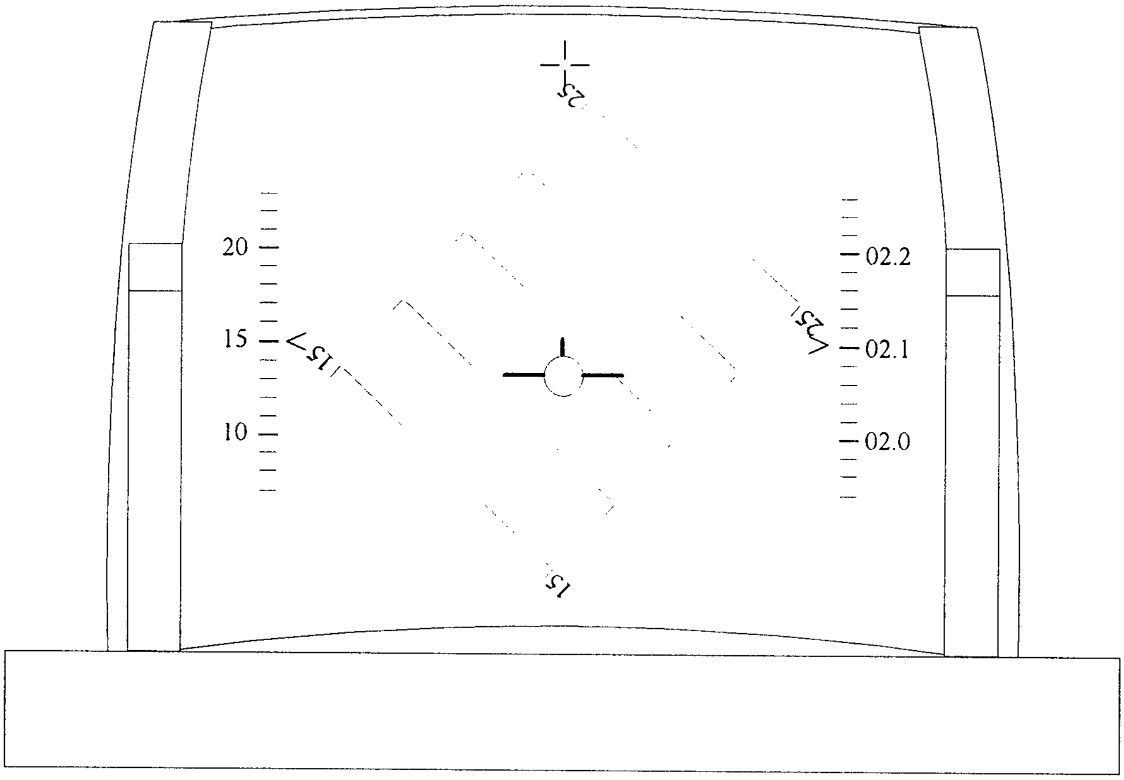

FIGURE 9. THE UNUSUAL ATTITUDES TEST.

4.3.8.3 Detailed Description of the Task.

As noted above, the basic stimulus display for this task will be the modified F-16 HUD display illustrated in Figure 9. The actual test administration can be viewed in several different contexts, determinable by the experimenter. For instance, one possible context would be that the subject would be in an isolated environment (i.e., one in that there was no previous stimulus conditions) and would suddenly see the HUD picture depicting an unusual environment. The subject's response to this condition would then be measured. In another situation, the subject might be engaged in a particular task (either an A-PASS test or a flight condition). The screen would then "blank" for some short period of time, indicating that the aircraft has entered into a cloud bank, or other condition in that the HUD information is unavailable. At the end of some short period of time, the HUD would reappear. However, there would be an unusual attitude indicated. Again, the pilot would have to take immediate action to correct this unusual attitude in order to avoid a catastrophic outcome.

The specific unusual attitudes to be employed in the APASS will be determined during Phase II. They will reflect the experience of our pilot consultants in employing this type of test, and will constitute a broad range of realistic unusual attitudes that can and frequently do occur in the experience of a normal pilot. Specifically, we would intend to incorporate at least 8 unusual attitude conditions. These would be roughly equated with respect to difficulty of the corrective action required. Again, the specific unusual attitudes to be employed will be determined during Phase II of this effort.

It would be anticipated that a single trial of this test will take between 30 seconds and 2.5 minutes to complete, depending on the timing of the initial unusual attitude and the duration of the corrective action. Unlike other tests, however, it would not be anticipated that this procedure would require a large number of trials in order to arrive at a reliable estimate of the pilot's capability. In fact, two to five trials would be sufficient.

On the other hand, it should be recognized that this procedure will require considerable training time in order to assure that the subjects are capable of performing the corrective action required. For a non-pilot subject, this training time might be as long as ten hours. The goal, of course, is to assure that the subject, under nominal conditions, is perfectly able to respond to the unusual attitude in an appropriate way. For some subjects, this training time could be less than one hour. Therefore, we estimate that the average training time for non-pilot subjects would be approximately ten hours. On the other hand, pilot subjects should already be well trained in responding to unusual attitudes. Therefore, for these subjects, one hour of training should represent a maximum of training time. In order to assure that they are proficient, however, a sufficient number of trials should be run to guarantee stability of performance. Thus, even in pilot subjects, at least 30 minutes of training time should be considered necessary.

The subject's response to the unusual attitude will, of course, be quite idiosyncratic. Not everyone corrects the unusual attitude in the same way. Therefore, no individual measures of performance (e.g., response time, accuracy, etc.) will be appropriate. Instead, the overall success or failure of the correction must constitute the basic metric of this procedure. This overall success or failure can be quantified in several ways. In the present case, the time to achieve stable flight parameters will be the basic variable. This will be assessed as a ratio relative to the time that expert pilot consider to be "optimal" for the required correction. In other words, if the expert pilots consider that the optimal correction could have been achieved in five seconds, and the subject achieved the correction in ten seconds, the ratio would be 2.0. This would then become the subject's score. Obviously, this metric is based on the assumption that the quicker the reaction (within the constraints of safety) the better it is.

Since this is a highly relevant operational measure, it will be necessary to have our pilot consultants explore other possibilities for operationally meaningful data from this task. However, as a minimum, it would be expected that the delay times or deltas uncovered by this test would serve as direct inputs into the computer models described below.

4.3.8.4 Potential Problems with this Procedure.

This essentially represents a fairly complex response pattern on the part of the subject. Whenever one encompasses such a complex response pattern, the danger is that the subject has many ways to respond to the stimulus situation. Therefore, the most significant potential problem with this procedure is in its interpretation. Given a particular unusual attitude, different experienced pilots may respond in different ways. Some of these might be as efficient (in terms of energy expenditure or shot advantage) as other optimal responses (in terms of time or absolute recovery).

It will be necessary, therefore, to assure that a subject is not penalized for carrying out an atypical but efficient maneuver. As a minimum, the solution to this problem is not to pay undue attention to the actual maneuver that is performed. In other words, considerable latitude must be allowed the subjects in how they recover. This means that the focus must be on the time taken to recovery and the final outcome.

In terms of the time taken to complete the maneuver, care must be taken to utilize, as far as possible, the subject as his or her own control. In other words, whenever possible, a baseline performance time for each type of maneuver should be established. This is necessary because of the difficulty of obtaining a "gold standard" with respect to the time that the maneuver could take from experienced pilots.

On the other hand, in experimental situations where it is impossible to use the subject as his or her own control, it should still be possible to utilize this task. This could be done either by developing a "standard" time through use of the computer modeling technique described below, or utilizing subject matter expert opinion as a last resort.

A second potential difficulty with this procedure is the amount of training time that will be necessary to assure subject proficiency. Most non-pilot subjects are not extremely familiar with the control actions that would be necessary to recover from an unusual attitude. It would therefore be anticipated that non-pilot subjects would require considerable training on this task (perhaps extending between 10 and 20 hours). In some applications, this training time could be prohibitive. In others, however, this may not be so serious a problem. In the case of acceleration research, subject panels typically have considerable amounts of time for training. In other situations, one might reduce the training time significantly by simply training the specific unusual attitudes that will be used in the experiment. In this way, depending on the number of attitudes to be used, training time could be reduced to one to four hours.

4.3.9 Test 9. Aircraft Recognition.

4.3.9.1 General Description of the Procedure.

It was suggested both by the Air Force performance meeting and by NTI pilot consultants that some measure of the individual's ability to recognize friend or foe (i.e., identify aircraft) be included in the A-PASS system. This was not given a high priority in the overall battery, but it was considered essential to provide a brief test of this critical aspect of the pilot's abilities. The simplest way to achieve this goal would be to use the standard "silhouette" recognition tests that have been used by the armed forces for many years. In this, well defined black and white silhouettes of aircraft would be flashed to the individual for a brief period of time, and recognition times and accuracies would be collected. Although this is a well-studied technique, it would introduce some challenges in extrapolating the results to computer mission models. Therefore, a variant of this traditional technique will be used in A-PASS.

In this A-PASS procedure, as in the standard one, black and white silhouettes of current aircraft will be used, and these will be presented for varying brief periods of time. However, in A-PASS the silhouettes will be shown in several different orientations, and at various sizes. The subject will have to decide whether the aircraft is "friend" or "foe", and then initiate an appropriate action as quickly as possible. The target will only remain on the screen for a specified period of time, and then will disappear. If the subject decides the aircraft is friendly (based on previous training) a simple button press will be required. However, if the aircraft is a foe, then he or she will be required to carry out pre-briefed control actions depending on the target's orientation.

4.3.9.2 Construct Analysis of this Task.

The basic desire here is to determine whether there is a decrement in the pilot's ability to quickly process overlearned visual symbols. The silhouette shapes that will be used will all be very familiar to the subject, but will not have been trained under a limited viewing time. Therefore, the shape presented briefly will have to be compared to a static memory of several similar shapes. This technique should tap elements of memory retrieval, as well as sensory registration. The requirement for the subject to respond to the identification and orientation adds a discriminative motor response to the entire sequence, making it more like what the pilot has to do in a real-world situation.

4.3.9.3 Detailed Description of the Task.

Specific details regarding the actual silhouettes to be used in this task will be determined during Phase II, in consultation with government and consulting pilots. The aircraft chosen will represent a reasonably exhaustive list of current friendly and enemy fighter aircraft. Each aircraft will be shown in four orientations: 1) horizontal and 90 degrees left, 2) horizontal and 90 degrees right, 3) vertical straight up, and 4) vertical straight down. All silhouettes will be high-contrast black and white.

This task may be administered either while the subject is performing another routine task (e.g., straight and level flight) or as a stand-alone test. In either case, the computer screen will "blank" for up to 250 milliseconds, and a silhouette will appear in the center of the screen (a dot or other fixation point may be supplied if necessary to maintain foveation). The silhouette will remain on the screen for one of ten durations, ranging from 150 milliseconds, up to 1.5 seconds. The screen will then become blank until the subject makes a response. At that point, a variable intertrial interval, ranging from 3 to 15 seconds, will occur, and the next trial will be given.

The subject's response to the silhouette will be pre-trained to the point of asymptote. However, the training will not be done under limited viewing-time conditions. In other words, in training, the subject will see a silhouette that will remain on the screen until the appropriate response is made. Training will be complete when no errors are made under these conditions over a block of trials that contains each silhouette at least twice. The following responses will be trained: 1) If the aircraft is a friendly fighter in any orientation, the subject should push the throttle forward, 2) If the aircraft is an unfriendly fighter, the subject should move the stick toward the nose of the threat aircraft. Obviously, during phase II, these responses will be tested empirically. If pilot consultants feel that other discriminative actions are more appropriate, they will be tested and, if validated, adopted.

Scoring on this task will essentially consist of reaction time and percent correct, and threshold sensitivity measures. However, it should be obvious that these can be categorized in many different ways: friendly vs. unfriendly, aircraft type, orientation, and any combination of these. In general, it would be unlikely that great differences in these various categories would occur, or that small differences would be important to the computer models. However, the analysis programs will permit retrieval of the data in any category. Of particular interest will be the threshold time of presentation that yields 50% correct recognition (the absolute threshold time of recognition). This value can serve as a good summary measure of the entire skill, and should prove extremely sensitive to stressors. Interpreted in terms of an absolute time decrement in the person's recognition threshold, this value should serve as a direct input into the computer models.

4.3.9.4 Potential Problems with this Task.

One of the primary possible problems with this procedure deals with the amount of training that each subject has. Non-pilot subjects presumably will be learning this somewhat esoteric task without previous experience. Pilots, on the other hand, would be expected to have developed some skill in this area, although this in itself could cause some problems due to negative transfer-of-training effects. It will be necessary, therefore, to assure that both groups reach criterion performance before the test is used. This will be especially important for the non-pilot subjects. They must demonstrate that they have reached such a high level of performance that there will be no question that they are as proficient as experienced pilots. This could take considerable training time, estimated to be five hours, with a range from two to eight hours.

4.3.10 Test 10. Flight Simulation Testing.

Although it was previously decided that the A-PASS System would be independent of a flight simulation, it is still felt desirable to incorporate some reasonably high-fidelity flight simulation into the overall system. Techniques to achieve this have been studied extensively, although with limited success. The current flight simulation being prepared by the Government for the Wright Patterson simulator provides a defensible approach to measurement of the pilot's generic capability.

Although there are certain limitations introduced because of the necessity of its operation in a closed-loop centrifuge environment, this technique still provides an attractive approach to evaluating the global effects of G forces on the ability of an individual to fly. As an ancillary technique, the current effort by NTI to develop a desk-top flight simulation of the F-16 aircraft might be used to achieve this purpose. During Phase II, these potential platforms for this task will be evaluated and one or both will be implemented.

Since the actual flight scenarios of interest to the Air Force will vary, it will not be possible to describe procedures for this task in detail. Similarly, some effort will be required in Phase II to define the metrics that will be produced by such simulations. However, the pilot consultants suggested that it would be desirable to have at least one "standard" mission that could serve as a constant data base. Several such critical missions were considered, and it was recommended that a take-off simulation could serve this purpose well, since the procedures are relatively standardized, and since the takeoff combines elements of manual control, procedural memory, and decision making (e.g., the V_1 decision point).

Obviously, many questions remain to be solved during Phase II concerning this task. However, the obvious face validity of the measures, as well as their ability to complement other measures of the A-PASS system in computer models, suggests that such an effort would be justified.

4.4 INTEGRATION OF COMPUTER MODELING TECHNIQUES.

The development of the A-PASS test battery will provide the Air Force with a capability to measure flight-relevant skills in an experimentally rigorous but practical way. However, it should be realized that the output of the synthetic tasks described above is still in traditional metrics, such as reaction time, distance error, percent correct, and accuracy. While these provide the most appropriate units of measurement for assessing specific skills in the individual, they should be recognized for what they are: reasonably pure measures of human performance skills that are independent of any particular performance environment or task demand. In other words, like linear measurements of inches or feet, these are generic metrics that then must be applied to a particular situation in order to have specific meaning. Unfortunately, in the vast majority of cases, the performance researcher has been content to report the generic measures (often with elaborate statistical analyses). It was left to the user to interpret these measures in terms of his or her performance and environmental requirement.

The problem has been that the tools for carrying out such extrapolation of generic data to specific tasks have not been readily available. Perhaps the closest that performance researchers have ever come to providing a scientifically defensible, quantified approach to this has been the control-theoretic analyses presented by individuals such as Jex, McDonnell, and Phatak (1966). In this, a subject's tracking behavior is quantified with extreme precision. Further, this tracking behavior is then related to closed-loop control requirements such as those encountered in an aircraft. Unfortunately, this powerful demonstration of the necessity of translating basic data into real-world results has a reasonably limited scope -- the closed-loop environment. For the vast quantity of performance requirements that are open-loop (e.g., those involving higher level cognitive processing) no attempt has been equally successful. As the tasks required of the pilot in high performance aircraft become more cognitive in nature, control-theoretic analysis (while still valuable) has become less useful.

In the absence of a generally accepted set of techniques for building a "bridge" between basic data and the real-world, two other approaches have found some degree of use. By far the most frequently used technique is the "common sense" approach. In this, a subject matter expert looks at the basic performance data and simply tries to estimate whether or not it will have a real operational impact. For instance, if a certain amount of decrement is found in an individual's reaction time, the subject matter experts (SME) attempt to determine whether that decrement is meaningful in various environments.

Although this expert-opinion approach has a considerable amount to recommend it in terms of employing individuals who should be able to make the extrapolation, it has not proven particularly useful in the majority of situations. First, it has been difficult to get SME's to agree on the operational impact of any given performance level. Second, SME's sometimes are subject to biases that cause them to either over- or under-estimate the subject population's abilities. Third, SME's frequently are reluctant to apply the laboratory data to the operational world at all because they feel that laboratory conditions are so artificial that they should not be extrapolated to the real world. Finally, (and related to all of the above) subjective extrapolations lack the scientific criteria of replicability and stability. They are therefore too easily dismissed by commanders or planners under operational pressures.

4.4.1 Network Models.

The second type of attempt to bridge the gap between laboratory data and the real world has utilized computer models to simulate the human's response to specific task conditions and requirements. This approach, that had its origin in the late 1950's and early 1960's, resulted in the development of the SAINT model, fundamentally developed for the U.S. Air Force. This model introduced the concept of "network modeling", as the term is being used here. A network can be defined as a series of interrelated activities that flow in a specific sequence. In its original conceptualization, task two followed task one's completion, and task three followed task two's completion in a fairly extensive, time-dependent sequence. Simplistically, one can understand this early model as yielding a total time taken to carry out a complex task. The total time was simply the summation of all the individual sub-task times. This "time required" was then viewed in relation to the "time available", and this ratio could be interpreted in a variety of ways to indicate constructs such as "workload" "task achievability" or "task slack time".

The SAINT model later incorporated a number of features that permitted it to model real world events more realistically. These included the ability to "branch" from one line of text to another dependent on some event during the task sequence, the ability to carry on parallel tasks, the ability to modify a task's performance probability based on the results of previous tasks, and the ability to inhibit initiation of some tasks until other tasks have been completed. Further development of the SAINT model was carried out, primarily under sponsorship of the U.S. Army, in that these and other sophisticated capabilities were added. In addition, the entire program was rewritten to be carried out on a PC. This development is currently marketed by Micro Analysis and Design, Incorporated as Micro Saint $_{TM}$. Other network modeling approaches were developed during the same period by other organizations.

The network modeling approaches described above have had considerable success in real world applications, especially those that are clearly time-dependent. However, the output of these models is not always easy to apply, especially if one is interested in complex missions that are not totally time-dependent. The result is that, while network models of complex human performance are recognized to have considerable merit and potential, they have not seen utilization in as wide a range of operational areas as they perhaps should.

4.4.2 Systems Models.

What is required in the present case is a step beyond the network modeling approach that will objectively translate output from this type of model into estimates or predictions of the actual <u>impact</u> of the pilot's performance capabilities across a broad spectrum of real missions. In effect, while a network model can use performance results taken from some experimental source, such as the A-PASS test battery, to generate a cohort of simulated control actions by the pilot, something must be used to translate these control actions into actual movement of the aircraft. In other words, <u>these control actions must be simulated in another computer model that actually contains the aerodynamics of the system of interest.</u>

One of the major innovations proposed for the Phase I effort was to explore the use of Government-developed "systems" models that might serve the function of converting data obtained from network models into system performance estimates. These systems models consist of extensive engineering and graphical representations of the response of given systems in the real world to particular control inputs. For instance, one model ("BLUE MAX III") simulates the movement of an aircraft in three dimensional space with defined control inputs. These movements are system specific. In other words, if one is utilizing an F-16 model within BLUE MAX, and specifies that the pilot (at a given altitude, speed and angle of attack) makes a hard-over jinking maneuver to the left at a given point in time, the model will calculate when and how far an actual F-16 aircraft would turn. In this way, one can discover the actual movement and momentum of an aircraft through space, dependent on what the pilot does at any given point in time.

4.4.3 Combining Network and Systems Models.

In Phase I, it was proposed that by utilizing data from the A-PASS test battery in a network model, one could determine the distribution of the pilot's command inputs for specific required maneuvers. Again, it should be noted that such a distribution is <u>system generic</u>. It simply says that after a given stressor the pilot, on average, may be X milliseconds slower, or X percentage less accurate in making a particular kind of command action, and that there is a specific distribution of performances around that average. It was proposed that by introducing these performance times and accuracy's into systems models, actual operational outcomes for a given mission could be determined.

These three elements (performance tests, network models, and system models) must be then integrated into one coherent, user-friendly package. In this way, the Government can: 1) quantify the performance effects of a given stressor (e.g., G-profile) on basic human performance, 2) translate those observed performances into descriptions of specific command actions by the pilot, and 3) determine the success or failure of that flight profile in a given weapon system. This could be done by the A-PASS test battery, a network modeling system such as Micro Saint, and appropriate mission models.

4.4.4 Phase I Demonstration.

During Phase I, a prototype system incorporating both a network model and a systems model was created and tested. One engineering systems model (BLUE MAX III) was obtained from the Survivability and Vulnerability Information Analysis Center (SURVIAC) at Wright Patterson Air Force Base and investigated for its potential to carry out the kind of data analysis proposed above.

The BLUE MAX program requires the user to specify the performance of the pilot (i.e., the command inputs) for each run of the simulation. This means that for each specific task in a given mission segment, it is necessary to supply it with a specific performance time or accuracy for every run of the BLUE MAX model (every "flight" of the aircraft). Further, these times and accuracy values must be expressed in terms of command inputs (e.g., where the aircraft will go, what the G-level will be, etc.). Fortunately, this is precisely the type of data that, with some modification, network models can produce.

Micro Saint, on the other hand, uses a monte carlo technique to make a probabilistic "decision" about the individual's performance time (and accuracy) at every node in the sequence (see "Accounting for interdependencies..." earlier in this report). This decision, of course, is recorded. In this way, Micro Saint, in effect, creates a single "flight". This flight represents a totally realistic picture of what one pilot might do on one flight, based on the distribution of what pilots usually do on such flights.

If the network model is run many times, a cohort of "flights" is produced that simulates what actual pilots would do over a large number (e.g., thousands) of flights. If one were to run the Micro Saint model of a particular engagement a thousand times (that could be done in a matter of hours) it would produce a time/accuracy history of each performance element in each and every mission. In effect, it would generate a thousand missions. Within these, there would be a few extremely poorly done missions, as well as a few missions that were exceptionally well done. However, the bulk would certainly fall in the middle, clustering around the "average" performance of the "average" pilot operating under certain conditions of stress.

Each of these missions could then serve as a single mission input to a systems model such as BLUE MAX. The systems model would be told, for instance, that on this particular flight, this individual was 75 milliseconds late in initiating a particular turn. BLUE MAX would then calculate where the aircraft would be X amount of time later as a result of that 75 millisecond delay. In that mission, the Micro Saint model may have picked an exceptionally fast time at another point in the mission. BLUE MAX would then calculate what that fast response time did to the aircraft's state vectors at X amount of time. Ultimately, of course, these state vectors of the aircraft will have an impact on the ultimate success of the mission.

The strategy that evolved in Phase I, then, was to utilize the A-PASS test battery data in Micro Saint models of particular mission segments. These data were used to generate a number of simulated flights. In this prototype demonstration, only a few of the many simulated flights produced by Micro Saint were run on the BLUE MAX model (since at this point the data entry still has to be done manually). The output of that model was used as input into an NTI-developed ballistic model, the output of that was used to determine the expected impact on the weapons delivery accuracy. This approach was demonstrated during Phase I, and is described below.

For the feasibility demonstration of this approach, NTI selected a pop-up, air-to-ground attack mission of an F-4E (although it should be noted that in the BLUE MAX model, many of the weapons systems -- bomb sight, weapons computer, INS, etc. -- appear to operate more like those in an F-16 aircraft). This aircraft selection was made purely due to the time constraints imposed by the Phase I schedule, and Government delays in obtaining some of the newer models. Other aircraft are also available within the BLUE MAX model. However, it would have taken some time to obtain them from SURVIAC. Similarly, other mission segments (e.g., air-to-air, close air support, refueling, etc.) could have been chosen. A reasonably small demonstration mission was selected for ease of presentation and understanding of the technique. The reader should be assured that this demonstration is truly representative of that that could have been done on other aircraft and missions.

The first step in the modeling exercise was to create the Micro Saint model. To achieve this, pilot consultants generated a pilot-oriented description of the tasks that must be done during the pop-up air-to-ground mission. The goal was to produce a simple (five to seven node) sequence of pilot initiated behaviors . The basic tasks selected by the pilots in the pop-up attack were: 1) an "offset maneuver", consisting of an initial 20 degree offset turn at an action point, 2) the pop-up maneuver itself, 3) a "pull-down" maneuver toward the target, 4) the roll-out to line up on the target, 5) a gunsight tracking of the target, and 6) a weapons release, or "pickle" point. For this initial demonstration, complex dependencies among these various nodes were not included.

It will be noted that each of these behaviors involve an action in that the pilot has to make a control input into the aircraft, in effect commanding the aircraft to achieve certain state vectors. It would be equally easy to use behaviors that simply modified some conditions in the aircraft, such as a radio frequency change. The essential requirement, however, is that the pilot must demonstrate a behavior that might eventually have an impact on the final mission. Therefore, "thinking about something" or choosing something is not an appropriate node in this analysis. Only when the operator does something in response to the thought or decision is a behavioral node created.

The simple Micro Saint model created in this exercise is presented in figure 10. This model also illustrates the relevant command outputs that the pilot must make at each node.

Obviously, the above activities from SME's are crucial to the validity of the entire product. In effect, the SME's are isolating the critical tasks that the pilot must perform, and that determine the quality of the final output of that performance.

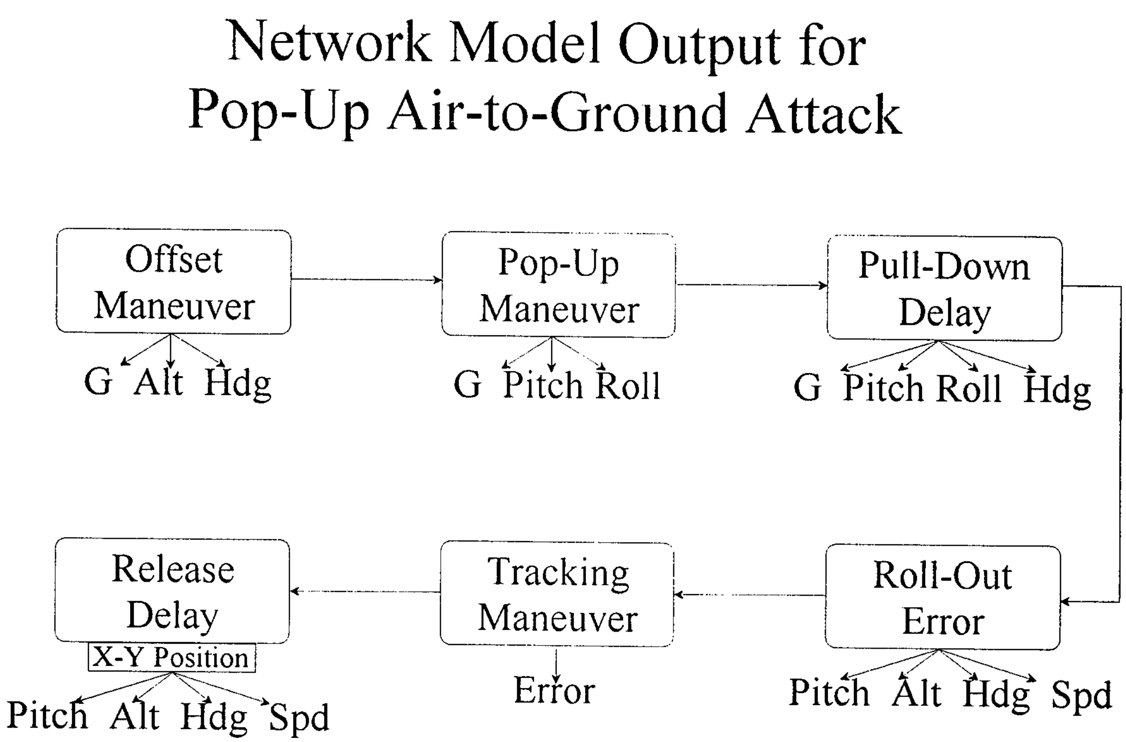

FIGURE 10. MICRO SAINT MODEL OF CRITICAL TASKS IN A POP-UP AIR-TO-GROUND ATTACK.

To the extent that this is done badly, or that critical elements in the mission segment are missed, the output will be less accurate. If most, or all, crucial tasks are included, and are appropriately modeled by the network model, mission outcome data will have an opportunity to be extremely precise.

The next step in the overall procedure is to assign performance values and distributions to the nodes in the Micro Saint model. It is at this point that the generality, relevance, and accuracy of the A-PASS test battery elements become crucial. As noted above, the A-PASS test battery was constructed to include as many critical flight tasks as could be anticipated. Therefore, it could be expected that an A-PASS test result will be found that is applicable to each of the nodes in any of the critical mission segments selected for a study. For instance, in the pop-up air to ground maneuver described above, there are elements of pitch and roll capture, gun tracking, and time inference. Data generated from subjects in- and out-of the centrifuge should be directly related to these elements of the pop-up maneuver. In other words, each node of the Micro Saint model will be described by a mean value and a distribution around that value. In fact, in constructing the nodes themselves, Micro Saint prompts the author to enter these values.

In the Phase I demonstration, hypothetical values were created for the relevant A-PASS tests (since the test battery does not yet actually exist). These values were selected by NTI performance specialists and pilot consultants to represent realistic estimates of how subjects will perform on the tests, both under nominal conditions and after a stress that caused some degree of performance decrement.

Having introduced nominal and distribution values into the Micro Saint model for each of the nodes, it was now possible to begin running the model itself. Ultimately, this will be a "turn-key" operation, in that one simply specifies the number of iterations to be performed. One might specify 1000 iterations. Each iteration would constitute one complete flight of the mission segment. Each iteration involves a monte carlo selection of performance characteristics at each node. These selections of course are stored, and constitute the output of the Micro Saint model. Once these values have been determined, the Micro Saint outputs are used in the systems model, as shown in figure 11.

4.4.5 Phase I Modeling Demonstration Results.

In this demonstration, only 37 "flights" were actually analyzed using the BLUE MAX model, since all data entry is still being done manually, and since the BLUE MAX program proved to be rather user-unfriendly. In the final system, of course, a program would permit automated transfer from the network model to the systems model, and extensive modifications would be made to the systems models to make them more user-friendly. These changes will permit analysis of an extremely large number of flights in a matter of minutes.

In order to demonstrate how the systems modeling technique will operate, three "missions" were selected from the total of 37, and the values at each point in the pop-up maneuver were manually entered into BLUE MAX. Several difficulties

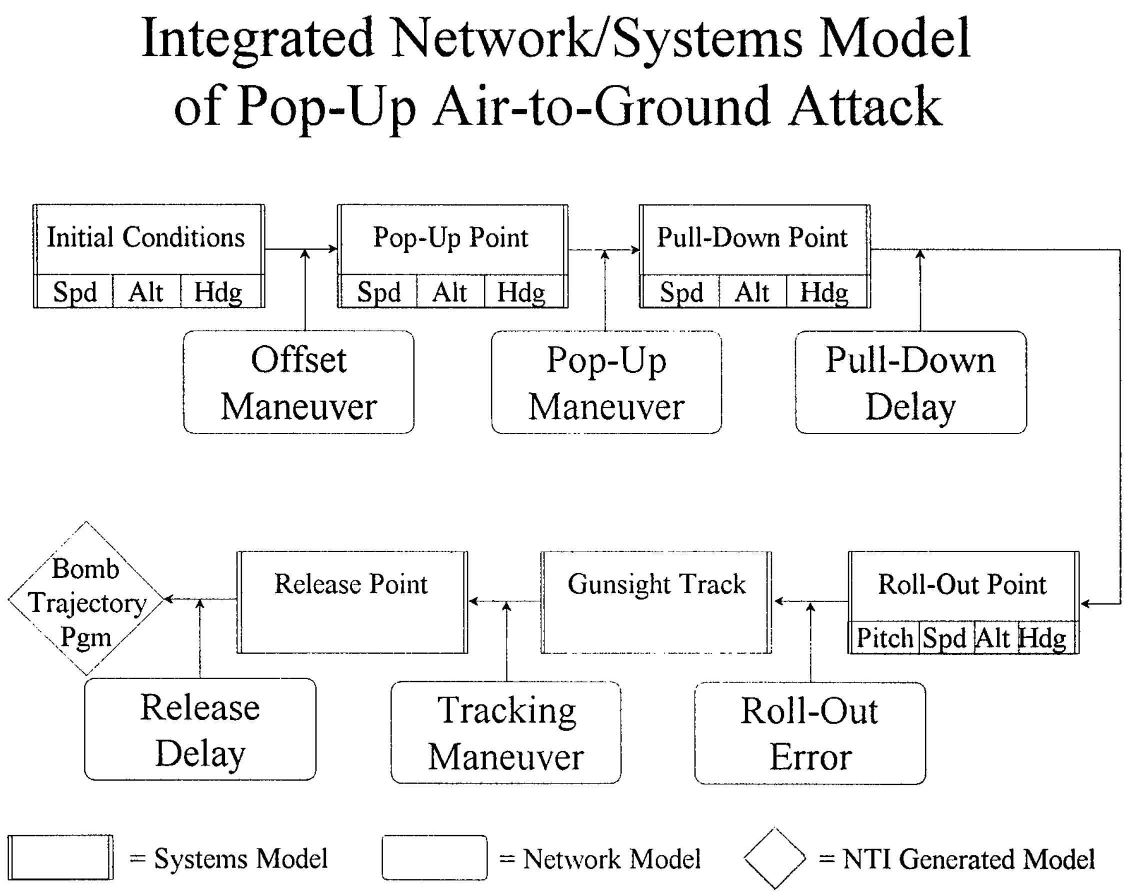

FIGURE 11. INTEGRATED MICRO SAINT/SYSTEMS MODELING OF THE POP-UP AIR-TO-GROUND MANEUVER.

were experienced with this operation, including apparent "bugs" in the program, and the fact that BLUE MAX has a limit of .1 second in any time estimates to be entered. Since performance differences are frequently smaller than this, BLUE MAX intrinsically does not have enough sensitivity for use in the way desired. However, while these types of problems were troublesome in this Phase I effort, modification of the program would be relatively straightforward, and would make the program entirely useable for the present purposes.

In spite of the above problems, the three "missions" selected revealed a wide range of "miss distances", or CEP's. In the flight closest to the overall "average", of all the flights run on Micro Saint, the miss distance was 26.1 feet. This would represent the standard error of the average undegraded pilot on this mission in the F4E aircraft. Selection of a set of values representing better performance resulted in a CEP of 17.9 feet, while selection of one of the poorest sets of performance data resulted in a CEP of 50.1 feet. These three missions therefore represent the kind of range of CEP values that one might obtain on a critical mission by a set of well-trained, undegraded pilots.

Finally, 34 "flights" typical of those that would be chosen by the Micro Saint model were "flown" on the BLUE MAX systems model. This model, taking the hypothetical values supplied from the Micro Saint flight, calculated the actual trajectory of the aircraft, given the control commands for each flight. In half of the flights (N = 17) errors in bomb release parameters reflected the distribution that would be achieved by a well-trained, proficient pilot. In the other half of the flights (N = 17), hypothetical values for a "degraded" pilot were used. A miss distance, given these parameters, was calculated for each flight.

These analyses revealed that, given this hypothetical type of data, the "nominal" condition would have produced a CEP of 25.0 ft. (SD = 10.8), while the degraded pilot would have produced a CEP of 50.1 ft (SD = 21.9). The range of CEP measures for the non-degraded pilots was 3.3 to 38.9 ft., while the range for the degraded pilots was 6.4 to 81.2 ft.

These data that, we continue to emphasize, are purely hypothetical, illustrate several points. First, of course, they show how the final output of A-PASS is phrased in terms that are immediately understandable to the operational commander. Second, they reveal that the "undegraded" pilots had a smaller dispersion in their performance than the degraded pilot. Finally, they illustrate that, using this approach, a few degraded pilots could end up performing as well as some non-degraded pilots. This reflects the reality of what actually happens in the real world better than the type of rigid mean values usually presented to the commander.

5. DISCUSSION OF POTENTIAL PROBLEMS.

It should be recognized that, even though each of the initial problem areas initially defined were addressed successfully in this Phase I effort, many questions and potentially thorny problems still exist. One of these deals with the translation of the critical tasks into performance resources or skills. This is essentially a theoretical exercise in that the performance specialist must analyze the tasks required of the pilot and determine that skills are involved. Although there are objective bases on that to make such inferences (e.g. factor analytic studies) there is not general agreement among performance specialists in how to perform this task. Therefore, different specialists might infer that different skills were required. In other words, a level of subjectivity is inherent in the whole APASS system. This is not necessarily a critical fault, since it should be possible to arrive at some general consensus of the skills involved. However, it does represent a potential source of disagreement with the results of the overall system.

A similar element of subjectivity is introduced in determining what test procedures best measure specific skills. Again, although there are experimental bases upon that such decisions can be made, there is no rigid set of rules for this translation.

The optimum solution to both of the difficulties noted above is procedural rather than scientific. Since the problem essentially involves vulnerability of the APASS system to criticism based on a lack of consensus in the field of performance assessment, the solution would be to obtain as much consensus as possible in the original formulation. In the Phase I demonstration above, time and facilities limited the number of consultants that could be employed to arrive at a consensus. However, in a final development of the APASS concept, the leading researchers in the field should be involved in making both of the interpretations; the taxonomy of skills, and the assessment of the constructs involved in the tests. This will not only assure the best scientific input, but will provide the highest level of scientific defensability for the overall approach.

Some technical problems also arose during Phase I in dealing with the network and systems models used. The particular network model used in Phase I was the Micro Saint model. This proved to be a relatively easy model to use, and to tailor into virtually any mission. However, it does not easily handle accuracy data, since it is essentially a time-based model. Some manipulation of the model is necessary to include accuracy or other kinds of non time-based inputs.

On the other hand, the systems models, as represented by the one used in Phase I, proved not to be user-friendly. Not only were data difficult to enter, but some of the limitations made the model as it exists much less useful for the present purposes than had been hoped. For instance, the smallest time increment permitted (.1 second) is not precise enough for most behavioral measures. It was also learned during Phase I that the problem of developing an interface that would allow the network model output to be used as input to the systems model is more complex than was originally understood. In the final APASS product, of course, it will be necessary to have the whole system integrated into one turn-key operation. This, it turns out, will require considerably more software development than had originally been estimated. All of these problems appear tractable especially within the capabilities of the Phase II development. However, they will require some re-assessment of the resources allocated to software development.

A similar and related difficulty that became apparent during Phase I was the overall complexity of the system, with related problems of training users and human-engineering the entire product. Although conceptually rather simple, the integration of all of the elements of the APASS approach clearly will require someone who is trained to operate the system, and preferably someone who understands the theoretical foundations of the system. Therefore, it appears unlikely that APASS in its present conceptualization would be useable in field situations by the actual operational commander, at least not without significant additional development and simplification of the process. Since APASS was originally conceived as a laboratory tool, this is not totally surprising. However, it would ultimately be desirable to define the entire procedure in the interest of simplification.

Even in view of the problem areas noted above, the path to actual development of APASS is reasonably clear. With minimal additional review and modification of the tests selected and the actual products to be used for modeling, the system is now ready to be built. One additional activity could profitably be interpolated. The APASS system designed during Phase I was specifically targeted to the acceleration environment. However, it should be recognized that this is a generic system. It will be applicable to any type of stressor that might affect performance. Therefore, it would be desirable to consider, during Phase II, any modifications that could be made to the system that would allow adaptation to other stressor environments or questions. These might include the heat chamber, the altitude chamber, questions of chemical defense effects, or even questions of the clinical utilization of drugs. Such adaptations would probably involve mostly hardware changes to adapt the apparatus to the extreme environment, however, in some cases, changes in the actual tests or interpretation might be considered.

6. REFERENCES.

Albery, W. B., Ward, S. L., & Gill, R. T. (1985). The effect of acceleration stress on human workload. (AAMRL-TR-85-039, DTIC-AD-A156770). Wright-Patterson AFB: OH. Armstrong Aerospace Medical research Laboratory.

Bennett, G. (1985). Aviation safety: Physical and mental incapacitations. Journal of Occupational and Environmental Health, 7 (Suppl.), 76-84.

Berkhout, J., O'Donnell, R. D., & Leverett, S. (1973). Changes in electroencephalogram spectra during repeated exposures to + Gz acceleration. (AAMRL-TR-72-123, DTIC-AD-A213927). Wright-Patterson AFB, OH: USAF Armstrong Aerospace Medical Research Laboratory.

Burton, R. R. (1986). G-induced loss of consciousness: Definition, history, and current status. Aviation, Space, and Environmental Medicine, 57, 2-5.

Chambers, R. M. & Hitchcock, L., Jr. (1963). Effects of acceleration on pilot performance (Tech. Report NADC-MA-6110) Johnsville, PA: US Naval Air Development Center (DTIC-AD: 408686).

Coburn, K. R. (1970). Physiological endpoints in acceleration research. Aerospace Medicine, 41, 5-11.

Englund, C., Reeves, D. L., Shingledecker, C. A., Thorne, D. R., Wilson, K. P., & Hegge, F. W. (1987). The Unified Tri-Service Cognitive Performance Assessment Program - UTCPAB: 1. Design and specification of the test battery. Report 87-10, U. S. Naval Health Research Center, P.O. Box 85122, San Diego, Cal., 92138.

Fong, K. L. (1992). Maximizing + Gz tolerance in pilots of high performance combat aircraft. (AL-SR-1993-0001). Wright-Patterson AFB: OH. Air Force Materiel Command.

Frankenhauser, M. (1945). Effects of prolonged gravitational stress on performance. Acta Psychologica, 104. 10-11.

Grether, W. F. (1971). Acceleration and human performance. (AFAMRL-TR-71-22). Wright-Patterson AFB: OH. Armstrong Aerospace Medical Research Laboratory.

Jex, H. R., McDonnell, J. D., & Phatak, A. V. (1966). A "critical" tracking task for man-machine research related to operator's effective delay time. Proceedings of the second annual NASA University Conference on Manual Control (Report No. NASA-SP-128). Cambridge, Mass.: Mass. Inst. of Technology, 361-377.

McCloskey, K. A., Tripp, L. D., Chelette, T. L., and Popper, S. E. (1992). Human Factors, 34(4), 409-428.

McMillan, G. R. Report of the Armstrong Laboratory Situation Awareness Integration (SAINT) team (briefing transcript). In Vidulich, M., Dominguez, C., Vogel, E., & McMillan, G. (1994). Situation awareness: Papers and annotated bibliography. Armstrong Laboratory, Wright-Patterson Air Force Base, Dayton, OH, 45432. (Rept. No. AL/CF-TR-1994-0085.)

O'Donnell, R. D. (1993). Enhancing soldier performance: I. A non-linear model of performance, selection testing, and training. Final Report, SBIR Phase I Contract 91ARI-015, Army Research Institute Contractor's Report.

O'Donnell, R. D., (1991). Directions in performance assessment: Results of the workshop on future metrics and models for assessment of human/system performance in advanced military systems. CSERIAC, Det. 1, AL/HE/CSERIAC, Wright-Patterson AFB, Ohio 45433-6573.

O'Donnell, R. D. and Eggemeier, F. T. (1986). Workload assessment methodology. In Boff, K. R., Kaufman, L. and Thomas, J. P. (Eds.) Handbook of Perception and Human Performance Vol II. Wiley, New York.

Perez, W. A. (1986). Performance metrics for use under sustained Gz acceleration: Review and recommendations. (Biodynamic Data Bank Accession No. 45997). Wright-Patterson AFB: OH. Armstrong Aerospace Medical Research Laboratory.

Perez, W. A., Masline, P. J., Ramsey, E. G., & Urban, K. E. (1987). Unified Tri-Services Cognitive Performance Assessment Battery: Review and Methodology. AAMRL-TR-87-007.

Sem-Jacobsen, C. W. (1961). Blackout and unconsciousness revealed by airborne testing of fighter pilots, Aerospace Med, 32 (1961) 247.

Shingledecker, C. A. (1990). Transitioning psychological methodologies to industrial human assessment technologies. Annual Meeting of the Technology Transfer Society, Dayton, Ohio, June 26-28.

Shingledecker, C. A. (1984). A task battery for applied human performance research. AFAMRL-TR-84-071. Air Force Aerospace Medical Research Laboratory, Wright Patterson Air Force Base, Ohio.

Tripp, L. D., & Albery, W. B. (1987). Development of an oxygen mask integrated arterial oxygen saturation monitoring system for pilot protection in advanced fighter aircraft. In Proceedings of the 25th Annual SAFE Symposium (pp. 84-88). Yoncalla, OR: Safety and Flight Equipment.

Von Gierke, H. E., McCloskey, K., and Albery, W. B. (1991). Military performance in sustained acceleration and vibration environments. In R. Gal and A. D. Mangelsdorff (Eds.) Handbook of Military Psychology. John Wiley and Sons: New York.

Weltin, M., Broach, D., Goldbach, K., and O'Donnell, R., (1992). Concurrent criterion-related validation of the air traffic control specialist Pre-Training Screen. (Final report delivered under FAA contract DTFA01-89-Y-01019). Washington, DC: Federal Aviation Administration Office of the Deputy Administrator.

Wickens, C. D. (1980). The structure of processing resources. In R. Nickerson & R. Pew (Eds.), Attention and Performance VIII. Hilsdale, N.J.: Erlbaum.

Wilson, G. F., Purvis, B., Skelly, J., Fullenkamp, P., and Davis, I. (1987). Physiological data used to measure pilot workload in actual flight and simulator conditions. Proceedings of the Human Factors Society.

U. S. Government Printing Office 1996 549-073/40023